ARCHITECTURE AND CORE OF THE SMALL RIBOSOMAL SUBUNIT

A Dissertation
Presented to
The Academic Faculty

by

Burak Gulen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology

May 2015

ARCHITECTURE AND CORE OF THE SMALL RIBOSOMAL SUBUNIT

Approved by:

Dr. Loren D. Williams, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Dev P. Arya
Department of Chemistry
Clemson University

Dr. Adegboyega K. Oyelere
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: March 19, 2015

To my parents.

ACKNOWLEDGEMENTS

I wish to deeply thank my advisor Dr. Loren Williams for his guidance, patience, and support. Without him it would not be possible for me to come this far.

I would like to thank my committee members Dr. Nicholas Hud, Dr. Adegboyega Oyelere, Dr. Roger Wartell, and Dr. Dev Arya for their wise inputs, suggestions and assistance during my research.

I also would like to thank Eric O'Neill, Dr. Anton Petrov, Dr. Chialong Hsiao, Dr. Brande Jones, and Jessica Bowman for helping me anytime I needed.

I would like to especially thank my parents, Figen and Hasan Gulen without whose guidance and unconditional support I would not be here.

I would like to present my special thanks to my special one, my soulmate, Susann Orth for her unconditional support. She was always there when I needed.

Lastly, I would like to thank every single person that helped me during my journey.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS IV

LIST OF TABLES IX

LIST OF FIGURES X

LIST OF SYMBOLS AND ABBREVIATIONS XIII

SUMMARY XIV

CHAPTER 1 INTRODUCTION 1

1.1 The Ribosome 1

1.2 Large Ribosomal Subunit and Peptidyl Transferase Center 3

1.3 Small Ribosomal Subunit and Decoding Process 4

1.4 Intersubunit Bridges 6

1.5 Transfer RNA Structure 6

1.6 Small Subunit Proteins 9

1.7 The Central pseudoknot 9

1.7.1 Motions within the small ribosomal subunit and the central pseudoknot 9

CHAPTER 2 LITERATURE REVIEW 12

2.1 Background 12

2.2 The Central pseudoknot is Essensial for Translation 12

2.2.1 Base complementarity of the helix 2 of 16S rRNA is essential for the ribosome function 16

2.3 The Central pseudoknot is Required for Stability of the Ribosome 16

2.4 Acylation of the Ribosomal Protein S5 is Affected by the Central pseudoknot 18

2.5 A Conformational Switch in Helix 27 of 16S rRNA 20

2.6 Arrangements of the Central pseudoknot Region of 16S rRNA 22

2.7 A Functional Relationship Between Helix 1 of the Central Pseudoknot and the Helix 27 Tetraloop 23

2.8 Conclusions 24

2.8.1 The contribution of this thesis on the small ribosomal subunit and the central pseudoknot 25

CHAPTER 3 REVISION OF THE SECONDARY STRUCTURE MAP OF THE SSU RNA 27

3.1 Introduction 27

3.1.1 Historical secondary structure maps of ribosomal RNAs 27

3.1.2 Relationship between 2D and 3D structures 28

3.2 Material and Methods 29

3.3 Results and Discussion 30

3.3.1 Looking at the central pseudoknot 33

3.4 Conclusion 35

CHAPTER 4 REVISION OF THE DOMAIN STRUCTURE OF THE SSU RNA 36

4.1 Introduction 36

4.2 Material and Methods 38

4.2.1 Superimposition 38

4.2.2 Shannon entropy 39

4.2.3 Domain criteria 39

4.3 Motivations for Research 40

4.3.1 Historical domain structure does not represent the 3D structure 40

4.4 Results and Discussion 41

4.4.1 Introducing domain A: Small ribosomal subunit domains radiate like helical spokes from domain A 41

4.4.2 Conservation of Domain A 42

4.4.3 Domain A shows tRNA mimicry 42

4.5 Conclusion 47

CHAPTER 5 ISOLATION OF DOMAIN A 48

5.1 Introduction 48

5.1.1 Defining the central core by a structural model 50

5.2 Material and Methods 50

5.2.1 Chemical reagents and synthetic oligonucleotides 50

5.2.2 Construction of the transcription vector for domain A^{ISO} RNA 50
5.2.3 *In vitro* transcription and purification of domain A^{ISO} RNA 52
5.2.4 Selective 2'OH acylation analyzed by primer extension reactions and dideoxy sequencing 53
5.2.5 Circular dichroism spectroscopy 55
5.2.6 3D modelling of domain A^{ISO} 55
5.2.7 Data mapping 56
5.2.8 Figures and images 56
5.3 Results and Discussion 56
5.3.1 *In vitro* folding of domain A 56
5.3.2 Helix 28 is an integral component of domain A 60
5.3.3 Disruptive mutations in the central pseudoknot prevent domain A^{ISO} folding to the native state 62
5.4 Conclusion 63
CHAPTER 6 INTERACTIONS OF DOMAIN A WITH RIBOSOMAL PROTEINS 65
6.1 Introduction 65
6.1.1 Tertiary binding proteins 65
6.2 Materials and Methods 66
6.2.1 Yeast-three hybrid assay 66
6.2.2 Electrophoretic mobility shift assay 70
6.2.3 *In vitro* folding studies 71
6.2.4 Preperation of the cloning vectors for purification of S5 and S12 72
6.2.5 Purification of S5 and S12 73
6.3 Results and Discussion 74
6.3.1 *In vitro* folding conditions for domain A RNA 75
6.3.2 *In vivo* interactions of ribosomal proteins analyzed by yeast-three hybrid system 78
6.3.3 *In vitro* interactions of ribosomal proteins analyzed by electrophoretic mobility shift assay 82
6.4 Conclusion 88
CHAPTER 7 EVOLUTIONARY MODEL OF THE CENTRAL PSEUDOKNOT AND SMALL RIBOSOMAL SUBUNIT 92

7.1 Introduction 92
7.1.1 Ribosomal Growth 92
7.1.2 Eukaryotic Expansions 94
7.1.3 Ancestral Expansion Segments 96
7.2 Materials and Methods 98
7.2.1 Secondary structures 98
7.2.2 Three-dimensional structures 98
7.3 Results and Discussions 98
7.3.1 Evolutionary Model for the Small Ribosomal Subunit 99
7.3.2 Evolutionary model for the central pseudoknot and the central core... 106
7.4 Conclusion 107
CHAPTER 8 CONCLUSIONS AND FUTURE DIRECTIONS 109
8.1 More accurate secondary structures 109
8.2 Revised domain structure 110
8.3 Domain A is confirmed as a domain 111
8.4 Protein S5 112
8.5 Evolutionary model for the SSU and the central pseudoknot 113
8.6 Proposed research for solving the central pseudoknot mystery... 114
8.6.1 *In vivo* SHAPE for screening the changes in the 16S rRNA structure with defected central pseudoknot 115
8.6.2 How to purify 16S rRNA and isolate the mutant rRNA from chromosomally enchoded ribosomes. 115
8.6.3 Cappilary electrophoresis and data processing 118
REFERENCES 119

LIST OF TABLES

Table 4.1 Pairwise superimposition of domain A RNA. 37

LIST OF FIGURES

Figure 1.1 X-Ray crystal structure of *Thermus thermophilus* ribosome. 2

Figure 1.2 Molecular interactions in the decoding center of SSU.....4

Figure 1.3 Intersubunit bridges are mapped on the secondary structures of 16S and 23S rRNAs. 7

Figure 1.4 Secondary and 3D structure of tRNA. 8

Figure 1.5 Binding locations of 30S ribosomal proteins to 16S rRNA.....10

Figure 2.1 Overview of special ribosome system by Brink and their results.....13

Figure 2.2 Close view of the central pseudoknot nucleotides.....14

Figure 2.3 30S initiation complex representation for Poot's study.....17

Figure 2.4 Conformational switch in the stem edge of helix 27 of SSU proposed by Dahlberg and Lodmell.....21

Figure 2.5 Functional relationship between helix 1 and helix 27 tetraloop.....24

Figure 3.1 Revised and the historical secondary structure of 16S rRNA. 31

Figure 3.2 A closer look at the central pseudoknot.....33

Figure 4.1 Revised domain structure of SSU rRNA from *T. thermophilus*. 38

Figure 4.2 Stacking interactions between helix 2 and helix 28 of domain A^{ISO} (and 16S rRNA). 41

Figure 4.3 Conservation of domain A structure. 43

Figure 4.4 Shannon entropy data is mapped on domain A rRNA secondary structure map.....44

Figure 4.5 Domain A mimics tRNA. 46

Figure 5.1 Secondary structure and three-dimensional model of domain A^{ISO}. .49

Figure 5.2 Secondary structure of domain A^{ISO} construct for SHAPE reactions...51

Figure 5.3 SHAPE reactivity of domain A^{ISO} mapped onto the proposed secondary structure in the presence of sodium and magnesium cations.....57

Figure 5.4 Comparison of SHAPE reactivity between domain A^{ISO} helix 27 and Weeks' intact 16S rRNA helix 27. 58

Figure 5.5 Circular dichroism spectroscopy of domain A^{ISO}. 59

Figure 5.6 SHAPE reactivity of the intact domain A^{ISO} and helix 28 excised mutant. 61

Figure 5.7 SHAPE reactivity of the intact domain A^{ISO} and C18A mutant.....62

Figure 6.1 Schematic representation of Y3H system for determining RNA-protein interactions. 69

Figure 6.2 Domain A DCTA and Mg^{2+} titrations in water. 75

Figure 6.3 Domain A DCTA and Mg^{2+} titrations in a buffer solution. 76

Figure 6.4 Domain A RNA folding in acidic and basic pH. 77

Figure 6.5 Yeast-three hybrid interactions between domain A and ribosomal protein S5. 79

Figure 6.6 Yeast-three hybrid interactions between domain A and ribosomal protein S12. 81

Figure 6.7 Electrophoretic mobility shift assay for domain A^{ISO} RNA and S5 fusion protein. 83

Figure 6.8 Electrophoretic mobility shift assay for P4-P6 RNA / 12 mer duplex RNA and S5 fusion protein.....85

Figure 6.9 SHAPE reactivity of the intact domain A^{ISO} and domain A^{ISO} ribosomal protein S5 fusion complex. 87

Figure 6.10 Circular dichroism spectroscopy of domain A^{ISO} RNA – ribosomal protein S5 complex. 89

Figure 7.1 Eukaryotic expansion segment es3 and its growth.....93

Figure 7.2 Insertion fingerprints for SSU rRNA growth.....97

Figure 7.3 Ancestral expansion segments of SSU rRNA and their temporal order. 100

Figure 7.4 Ancestral intersubunit bridges B3, B2b, and B2c. 101

Figure 7.5 Phases of the SSU rRNA evolutionary model.....104

Figure 7.6 Origin and evolution of the central pseudoknot and decoding center of the SSU. 105

Figure 8.1 Modifications in helix 45 of E. coli 16S rRNA. 117

LIST OF SYMBOLS AND ABBREVIATIONS

1M7	1-methyl-7nitroisatoic anhydride
2D	Two-dimensional
3-AT	3-amino-1,2,4-triazole
3D	Three-dimensional
APS	Ammonium persulfate
CAT	Chloramphenicol acetyl transferase
CD	Circular dichroism
cDNA	Complementary DNA
Cryo-EM	Cryo-electron microscopy
DCTA	Diaminocyclohexanetetraacetic acid
ddGTP	Dideoxy guanosine triphosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
Domain A^{ISO}	Isolated domain A
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
f-met tRNA	Formylmethionine tRNA
FR3D	Find RNA 3D
LSU	Large subunit of ribosome
MBP	Maltose binding protein
mRNA	Messenger RNA
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEI	Polyethylene imine
pI	Isoelectronic point
PTC	Peptidyl transferase center
RMSD	Root-mean square deviation
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
rRNA	Ribosomal RNA
s^4U	4-thiouridine
SDS	Shine-Dalgarno sequence
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SHAPE	Selective 2'OH acylation analyzed by primer extension
SSU	Small subunit of ribosome
TEMED	N,N,N,N-tetramethyl-ethylenediamine
tRNA	Transfer RNA
rProteins	Ribosomal proteins

SUMMARY

This thesis presents my work on the small ribosomal subunit, its architecture, and most importantly a previously unobserved domain in the heart of the small subunit of ribosome.

The ribosome is a highly significant molecular machine of all living species, synthesizing proteins. While there is no doubt that understanding of the translational machinery has advanced rapidly in the last decade, there are still many aspects of translation that remain unclear. Areas of active research include translocation mechanisms, dynamic molecular movements during translation, and the origins of the ribosomal subunits.

To increase the understanding of the small ribosomal subunit, we have developed structure-based secondary structures for ribosomal RNA for all domains of life. Our approach supersedes co-variation analysis based secondary structures and provides more accurate secondary structures.

Furthermore, we have developed a new domain definition for large RNAs. This definition is based on our observations of structure and interactions in the small subunit. Besides the lack of well-defined boundaries between domains, historical definitions do not reflect the three-dimensional structure of the small subunit accurately. We define a domain by self-consistent molecular interactions and an ability to fold in native state independent from the surrounding RNA. This definition reveals a new core domain (domain A) for the small ribosomal subunit RNA, from which the other domains radiate. Domain A contains the central pseudoknot. We have characterized and confirmed the integrity of domain A as

a domain by a variety of methods such as selective 2'OH acylation analyzed by primer extension, divalent cation dependency, circular dichroism, mutations and truncations, and protein interactions. Another interesting aspect of domain A we have observed is structural mimicry of tRNA. This mimicry suggests duplication events early in the history of the translational machinery, with implications for origins for genetic code.

Moreover, we have found destabilization of domain A and the central pseudoknot of the small ribosomal subunit upon binding of excess ribosomal protein S5. Since ribosomal proteins S5 and S12 interact with domain A of the small subunit of ribosome in native structure in 1:1 stoichiometry, our results support previously proposed 30S subunit assembly and biogenesis pathways in that ribosomal proteins S5 and S12 are tertiary binding proteins and therefore can only bind upon binding of primary and secondary proteins.

Additionally, we have proposed an evolutionary pathway for the small subunit of ribosome through ancestral expansion segments and insertion fingerprints. Our evolutionary model reveals the origins of the central pseudoknot and domain A of the small ribosomal subunit.

CHAPTER 1

INTRODUCTION

Ribonucleic acid (RNA) is a special macromolecule involved in life's most fundamental processes. A hydroxyl (OH) group at the 2' position of the ribose distinguishes RNA from deoxyribonucleic acid (DNA), giving RNA distinctive folding properties and chemical reactivity.

RNA can both encode information, as seen in RNA viruses (Ada & Perry, 1954; Chao & Schachman, 1956) and catalyze chemical transformations, as seen in ribozymes (Doudna & Cech, 2002; Lund et al., 2004). When it comes to the most significant and universal function of RNA, we have to elaborate the gigantic ribonucleoprotein (RNP) complex, the ribosome.

The protein synthesis machinery is highly complex and tightly regulated, and is one of the most vital processes in cells. Rapidly growing cells dedicate 60% of the all transcriptions to ribosomal RNA (rRNA) production. Fifty percent of RNA polymerases and 90% of messenger RNA (mRNA) splicing are serving to produce ribosomal proteins (rProteins) (Warner, 1999). To this end, understanding life itself is directly correlated to understanding the ribosome and its function.

In this chapter, we will take a closer look at the ribosome and its fundamentals.

1.1 The Ribosome

The ribosome is a complex molecular machine that synthesizes proteins in all living organisms. The ribosome has two subunits. The large ribosomal subunit

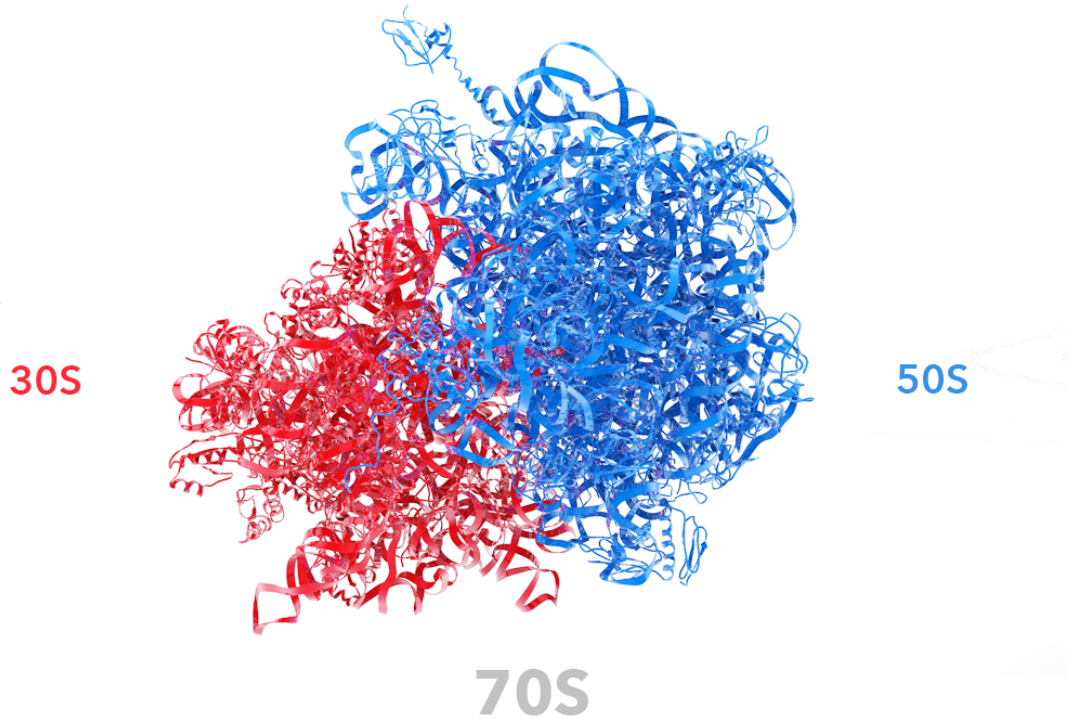

Figure 1.1 X-Ray crystal structure of *Thermus thermophilus* ribosome.
SSU is red, LSU is blue. Coupled 70S ribosomal unit is complexed with tRNAs and mRNA. (PDB ID: 2j00, 2j01)

(LSU) contains the peptidyl transferase center (PTC), which carries out the peptide bound formation during protein synthesis. The small ribosomal subunit (SSU), which contains the decoding center, provides the suitable molecular atmosphere for decoding the mRNA. Briefly, prokaryotic ribosome has 23S rRNA, 5S rRNA, and 35 rProteins in the 50S LSU; and the 16S rRNA and 21 rProteins in the 30S SSU. The associated subunits constitute the 70S bacterial ribosomal particle (Selmer et al., 2006). Eukaryotic ribosome is larger and in yeast is composed of the 40S SSU and the 60S LSU, which assemble to form the 80S eukaryotic ribosome.

1.2 Large Ribosomal Subunit and Peptidyl Transferase Center

The LSU contains the catalytic functionality of the ribosome. The catalytic core of the LSU is called the peptidyl transferase center (PTC). The PTC is conserved in sequence and in structure across the phylogenetic tree. The bacterial LSU RNA has around 2900 nucleotides within the 23S rRNA, which contains approximately 100 helices. Additionally, it contains the 5S rRNA, which is 120 nucleotides and separated to 5 helices. The LSU RNA forms a large assembly with 35 ribosomal proteins to form the 50S subunit. LSU RNA was previously ascribed to 6 domains as domain I through domain VI (Noller et al., 1981). However, recently Petrov and coworkers proposed a 7 domain structure for LSU RNA and introduced domain 0 as the structural center for 23S rRNA (Petrov et al., 2013). PCT is located in domain V of the 23S rRNA.

There are several proposed mechanisms for the reaction catalyzed by PTC (Muth, Ortoleva-Donnelly, & Strobel, 2000; Nissen, Hansen, Ban, Moore, & Steitz, 2000; Schmeing, Huang, Strobel, & Steitz, 2005). However, those mechanisms have been disproved experimentally. Universally conserved adenosine in 2451 nucleotide position (A2451) is suggested as a catalytic residue for those mechanisms. However, mutation experiments reveal that PTC catalyzes transpeptidation even without the A2451 (Thompson et al., 2001). While the chemistry behind the transpeptidation in PTC is not well explained, Rodnina and coworkers suggested the reaction catalyzed by PTC as a result of an entropy trap (Sievers, Beringer, Rodnina, & Wolfenden, 2004). The entropy trap mechanism suggests that the ribosome is catalyzing the peptide bond formation by stabilizing the substrates in the right place, orientation and the time.

1.3 Small Ribosomal Subunit and Decoding Process

The SSU plays the major role in maintaining the genetic code, and correctly translating the mRNA. The functional core of the SSU is located in the 3' minor domain, more specifically in the region of helix 44, which is called the decoding center. The bacterial SSU has approximately 1500 nucleotides, which forms the 16S rRNA. In addition to the rRNA, 21 ribosomal proteins stabilize and constitute the 30S ribonucleoprotein complex. The 16S rRNA contains 45 helices and has traditionally been ascribed to four different domains, which are the

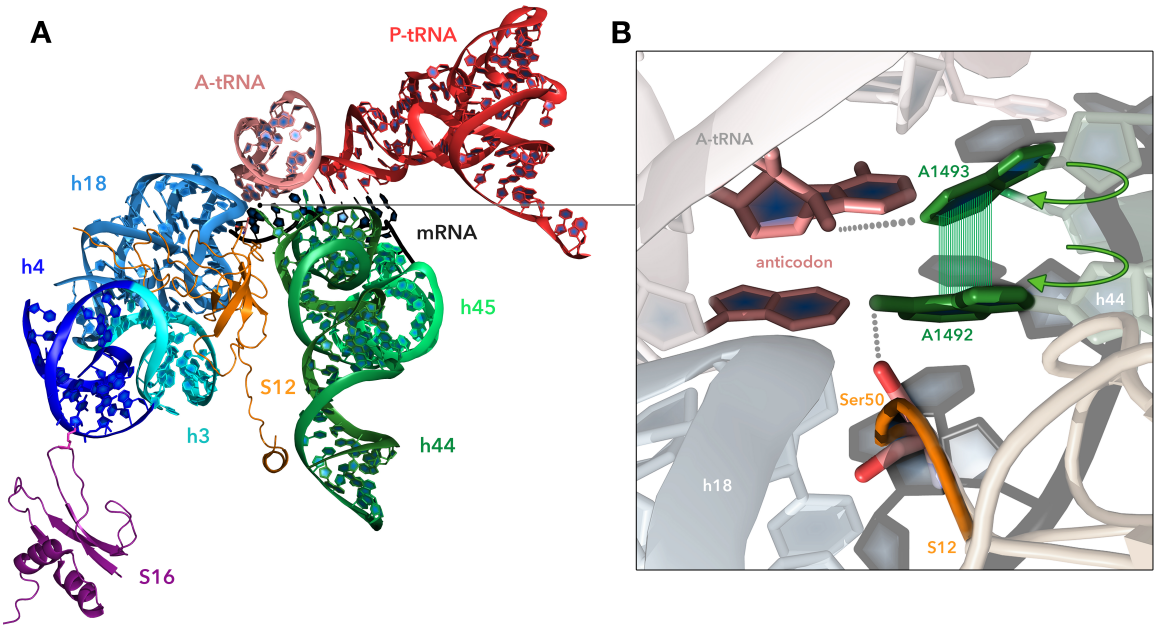

Figure 1.2 Molecular interactions in the decoding center of SSU.
(A) Overview of decoding site. Ribosomal protein S16 (purple) stabilizes helix 4 (dark blue) and causes a conformational change in helix 3 (cyan). The pseudoknot at helix 18 (blue) is stabilized by helix 3. Ribosomal protein S12 (orange) stabilizes the helix 18 and helix 44 (green) A-site adenosines. Codon-anticodon interactions of mRNA (black) and tRNAs from A-site (salmon) and P-site (red) are also seen. (B) Closeup view of decoding interactions. Universally conserved A1492 and A1493 flip out from helix 44 during decoding process (Green arrows illustrate flipping out). Serine 50 of ribosomal protein S12 stabilizes A1492 via H-bonding interaction. Stacking interaction between A1492 and A1493 is illustrated in the figure as green lines. A1493 involves in stabilization of codon-anticodon interaction between A-site tRNA and mRNA via H-bonding.
(PDB ID: 2j00) (Selmer et al., 2006)

5'domain, central domain, 3' major domain, and 3' minor domain. We recently revised the domain architecture of SSU and showed that the four domains of SSU radiate from a central core domain. Our revision of the domain structure of SSU is discussed in chapter 4 .

Helix 44 governs the decoding center of SSU (Figure 1.2). Helix 18 in the body region of 16S rRNA has a critical role in the decoding center as well. Helix 18 with its bulge and loop regions form a pseudoknot structure which facilitates A-site tRNA binding to the decoding center and is essential for ribosome function (Powers & Noller, 1991). Ribosomal protein S16 binds to helix 4 in the 5' domain of SSU and causes a conformational change in helix 3 (Ramaswamy & Woodson, 2009). Further, helix 3 stabilizes the pseudoknot located in helix 18. Ribosomal protein S12 both binds the helix 18 pseudoknot and A-site bulge of helix 44 as well as the central core region of SSU (Selmer et al., 2006). S12 stabilizes the universally conserved nucleotide A1492 via hydrogen bonding (H-bonding) that has further stacking interactions with universally conserved A1493. Those adenosines can assume two different conformations. During recognition of the cognate tRNA, those adenines flip out from helix 44 through codon-anticodon site of decoding process (Ogle et al., 2001). Nucleotide A1493 stabilizes 2' OH group of ribose backbone of anticodon loop of A-site tRNA. All those interactions seem crucial for translation machinery and are illustrated in Figure 1.2.

In prokaryotes the SSU also governs the anti-Shine-Dalgarno sequence (anti-SDS) near 3' end of the 16S rRNA. The anti-SDS is complementary to Shine-Dalgarno sequence on the mRNA, which binds to the SSU during translation initiation (Shine & Dalgarno, 1974). While decoding of mRNA occurs in the

decoding center of SSU, amino acids bond to the CCA terminal of tRNA go transpeptidation reaction catalyzed in the PTC.

1.4 Intersubunit Bridges

SSU and LSU form the 70S complex in bacterial ribosome during protein synthesis. There are 12 intersubunit bridges between ribosomal subunits that form stable and dynamic connection during translation. Intersubunit bridges were first observed by cryo electron microscopy (cryo-EM) (Lata et al., 1996) and further elaborated with increasing atomic resolution of the ribosomal structures (Gao et al., 2003; Gao & Frank, 2006; Noeske & Cate, 2012).

Specific locations of the intersubunit bridges are illustrated in Figure 1.3. In the SSU, most of the intersubunit bridges are located in 3' minor domain while domain IV governs most of the bridges in LSU.

1.5 Transfer RNA Structure

Transfer RNA is the amino acid carrier macromolecule in the cell. Francis Crick suggested an adapter molecule in protein synthesis after his discovery of double helix structure of DNA together with James Watson. Robert Holley proposed primary and secondary structures for tRNA in 1965 (Holley et al., 1965). In early 1970s the first crystal structures of tRNA appeared (Kim et al., 1973; Ladner et al., 1975).

Transfer RNA has cloverleaf secondary structure. In three-dimensions, coaxial stacking interactions and hydrogen bonding lead tRNA to fold in "L-shape" geometry (Quigley & Rich, 1976; Rich & RajBhandary, 1976). As for dividing the tRNA structure to subsets, tRNA contains the amino acid acceptor

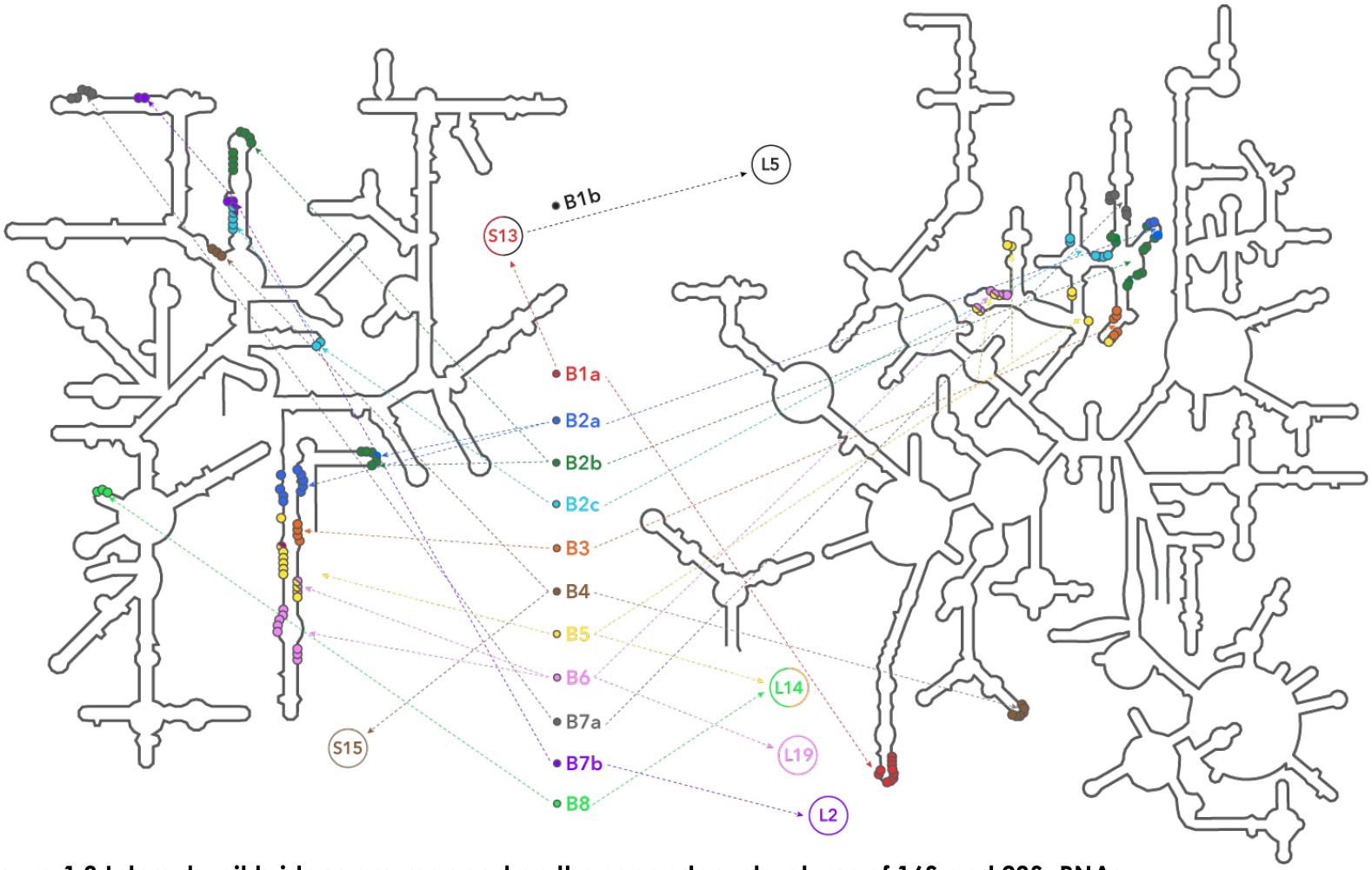

Figure 1.3 Intersubunit bridges are mapped on the secondary structures of 16S and 23S rRNAs.
Twelve intersubunit bridges are illustrated with different colors. Ribosomal preteins which involve intersubunit bridge interactions are shown in circles with the color of the corresponding bridges. Bridge locations are based on the work of Cate and coworkers.

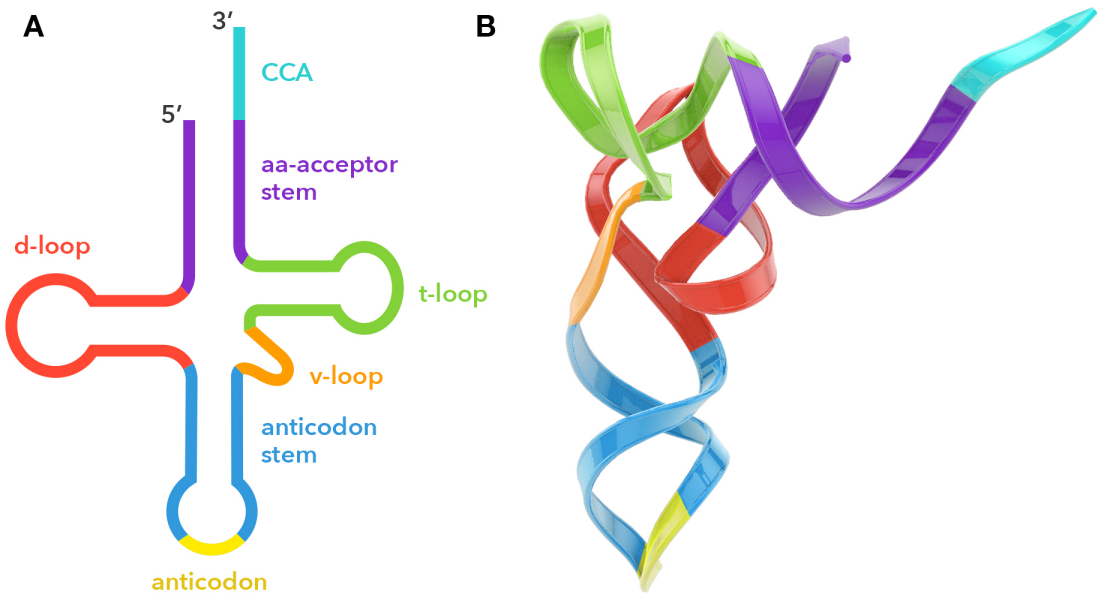

Figure 1.4 Secondary and 3D structure of tRNA.
(A) Secondary structure of tRNA in cloverleaf shape colored as anticodon (yelow), anticodon stem (blue), D-loop (red), V-loop (orange), T-loop (green), amino acid acceptor stem (purple), and CCA end (cyan). (B) X-ray crystal structure of P-tRNA^{Phe} from *Thermus thermophilus.* (PDB ID: 2J00)

stem, CCA tail, T-Stem, T-loop, D-stem, D-loop, V-loop, and anticodon stem and anticodon loop (Figure 1.4). The anticodon triplet sequence forms base pairs with mRNA codon during translation and maintains the correct selection of amino acids in SSU. Meanwhile, an amino acid, which is covalently attached to CCA tail of tRNA from it 3'OH group by an enzyme called aminoacyl tRNA synthesize, undergoes peptide bond formation in the PTC of the LSU. This mechanism explains the significance of "L-shape" geometry of tRNA, which requires interacting with mRNA in the decoding center of the SSU and also the PTC of the LSU, which are separated by 70 Å.

1.6 Small Subunit Proteins

There are 21 ribosomal proteins that bind to 16S rRNA in *E. coli* (Wilson & Nierhaus, 2009). Proteins are named by their individual sizes. S1 has the largest molecular weight while S21 has the smallest (Stelzl, Connell, Nierhaus, & Wittmann-Liebold, 2001). Nomura and Mizushima showed that protein binding to ribosomal RNA is hierarchical (Mizushima & Nomura, 1970). There are primary binding proteins such as S4, S7, S8, S11, S15, S17, and S20, which do not require additional proteins to bind the 16S rRNA. Furthermore, secondary binding proteins such as S6, S9, S13, S16, S18, and S19 call for primary binding proteins to bind the 16S rRNA first. Additionally, S2, S3, S5, S10, S12, S14, and S21 are tertiary binding proteins, which bind to 16S rRNA after primary and secondary binding proteins bound (Bunner, Beck, & Williamson, 2010; Mulder et al., 2010) (Figure 1.5).

1.7 The Central pseudoknot

The central pseudoknot of the SSU is first predicted by Pleij and coworkers (Pleij, Rietveld, & Bosch, 1985). The central pseudoknot plays a critical role in ribosome structure. It is located in the core of the SSU and is vital for ribosome function (Brink, Verbeet, & De Boer, 1993).

1.7.1 Motions within the small ribosomal subunit and the central pseudoknot

There are some molecular movements around the central pseudoknot that allow ribosome to mold its conformation during translation. The motions are directly correlated with mRNA and tRNA translocation. Two subunits are rotated in a ratchet movement, allowing the triplet code selection during decoding

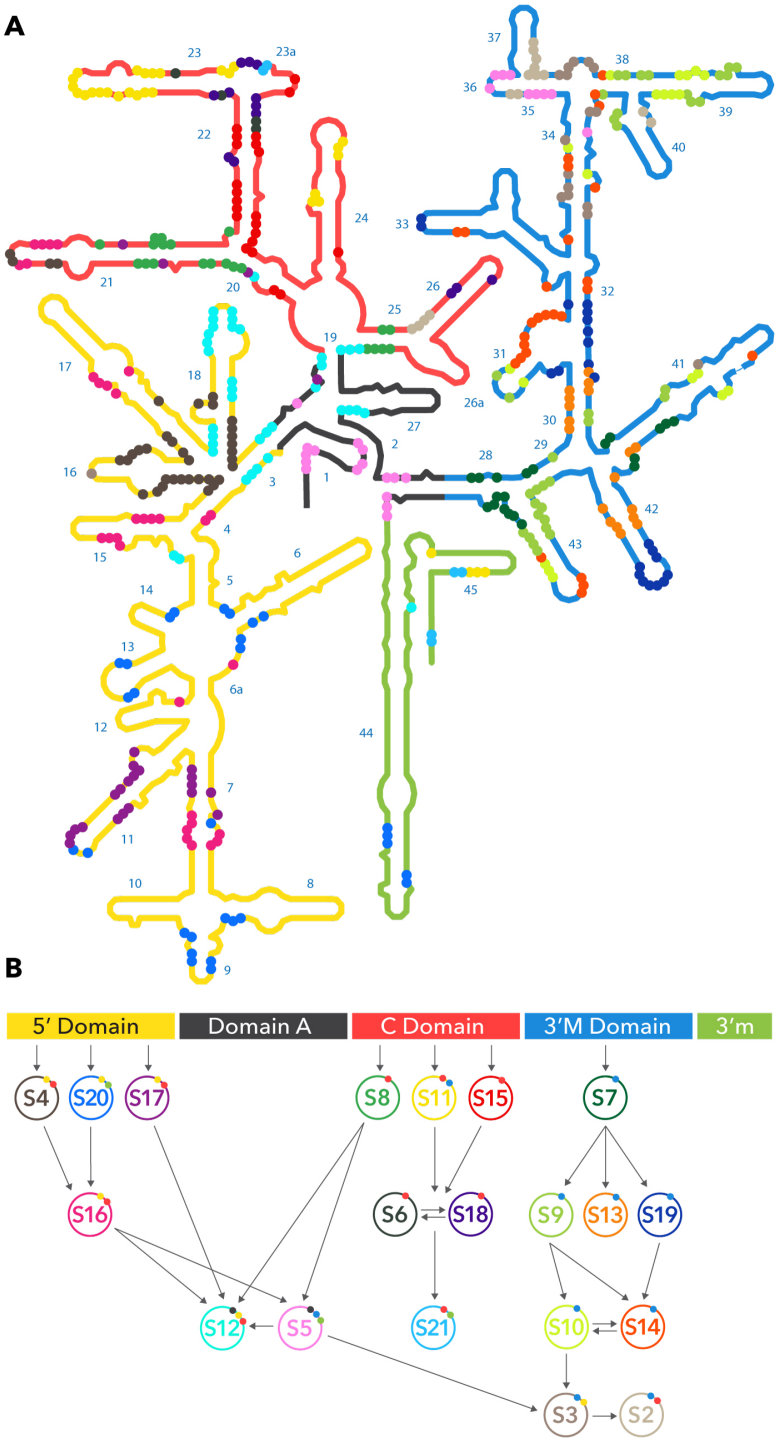
A
B
5′ Domain
Domain A
C Domain
3′M Domain
3′m
S4
S20
S17
S8
S11
S15
S7
S16
S6
S18
S9
S13
S19
S12
S5
S21
S10
S14
S3
S2

Figure 1.5 Binding locations of 30S ribosomal proteins to 16S rRNA.
(A) The 21 SSU proteins binding locations mapped on to the secondary structure of *Escherichia coli* 16S rRNA. Each protein represented with a different color circles. SSU domains are illustrated as contur lines with different colors. 5' domain (yellow), domain A (black), central domain (red), 3' major domain (blue), 3' minor domain (green). Helix numbers are shown in the secondary structure. Colors for each individual protein are seen on the panel B. Data for protein binding locations are obtained from ribovision server. (B) Assembly map of SSU proteins based on the study of Williamson and coworkers. Each protein represented as large circles. The SSU domains that ribosomal proteins bind is illustrated as mini filled circle on the corner of each protein representation. For example, S4 primarly binds to 5' domain and also binds to central domain (yellow and red filled circles indicate those domains). (This figure legend belongs to the figure in the previous page)

process (Guo & Noller, 2012). The role of the central pseudoknot in translocation is critical for the ribosome function. A detailed review of the central pseudoknot is given in chapter 2.

CHAPTER 2

LITERATURE REVIEW

The small ribosomal subunit plays a crucial role in decoding the mRNA during translation. Translocation of tRNA and mRNA allows the ribosome to add nascent amino acid to the peptide sequence. The mechanism of translocation is based on complex molecular movements within the rRNA and yet barely understood. To gain better insight into the molecular pathways of biochemistry, researchers employ a variety of techniques to solve the puzzle of translation.

2.1 Background

Before the high-resolution X-ray crystal structures era arose, visualization of nucleotide resolution of the life's most central machinery was very limited. Even with no 3D structures available for the small ribosomal subunit, the prediction of the central pseudoknot of 16S rRNA by Pleij and coworkers brought the structure and function of the pseudoknot to focus (Pleij et al., 1985). Later studies confirmed integrity of the central pseudoknot and its importance in ribosome function (Brink, Verbeet, & De Boer, 1993).

2.2 The Central pseudoknot is Essensial for Translation

Brink and coworkers tested the significance of the central pseudoknot by introducing disruptive mutations. They mutated the C18 and the G917 positions of the central pseudoknot and observed *in vivo* translation activity by using a specialized ribosome system in *E.coli* (Brink, Verbeet, & De Boer, 1993). A close view of those nucleotides is shown in Figure 2.2.

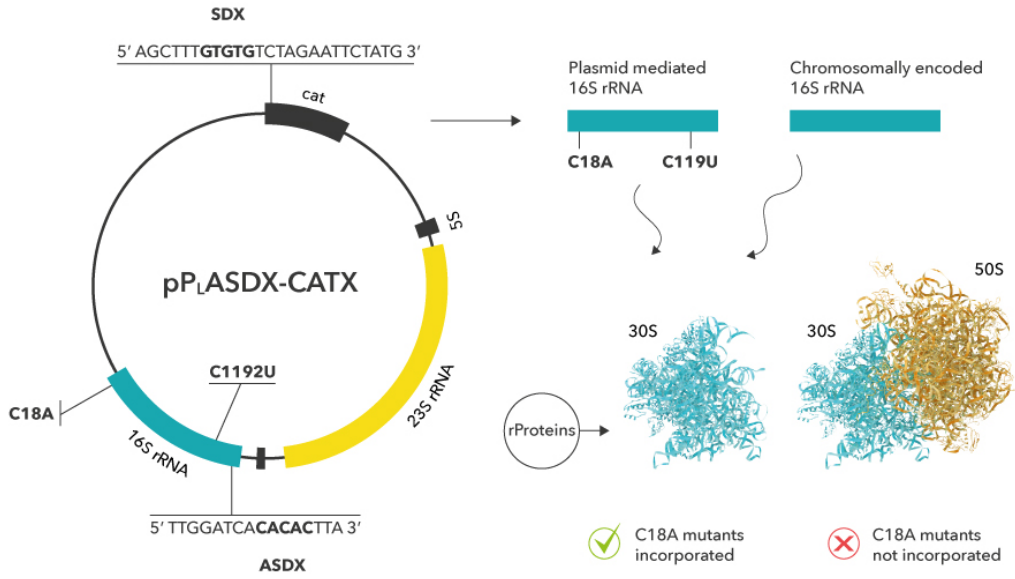

Figure 2.1 Overview of special ribosome system by Brink and their results. Plasmid pP$_L$ASDX-CATX is shown on the left. Anti-Shine-Dalgarno (ASD) sequence of the 16S is modified to compliment the Shine-Dalgarno (SD) sequence of the protein CAT mRNA. Results of the C18A mutant ribosomal RNA summarized on the right side of the figure. Mutant rRNA governing disrupted central pseudonot was incorporated in 30S subunits however not in 70S tight couples.

They altered the anti-Shine-Dalgarno sequence of 16S rRNA and used a complimentary Shine-Dalgarno sequence for the mRNA by using plasmid pP$_L$ASDX-CATX to monitor translation activity of the desired protein without interference by chromosomal ribosomes in *E. coli.* This plasmid contains a copy of one ribosomal RNA operon (*rrnB*). The rrnB operon in this plasmid governs mutations in the anti-Shine-Dalgarno sequence of the 16S rRNA as well as mutations in the Shine-Dalgarno sequence of the mRNA of the protein chloramphenicol acetyl transferase (CAT), which only specialized ribosomes are capable of translating (Figure 2.1).

Furthermore, to monitor the protein CAT, they did not only observe the activity of the protein by measuring the concentration of the product [^{3}H]-

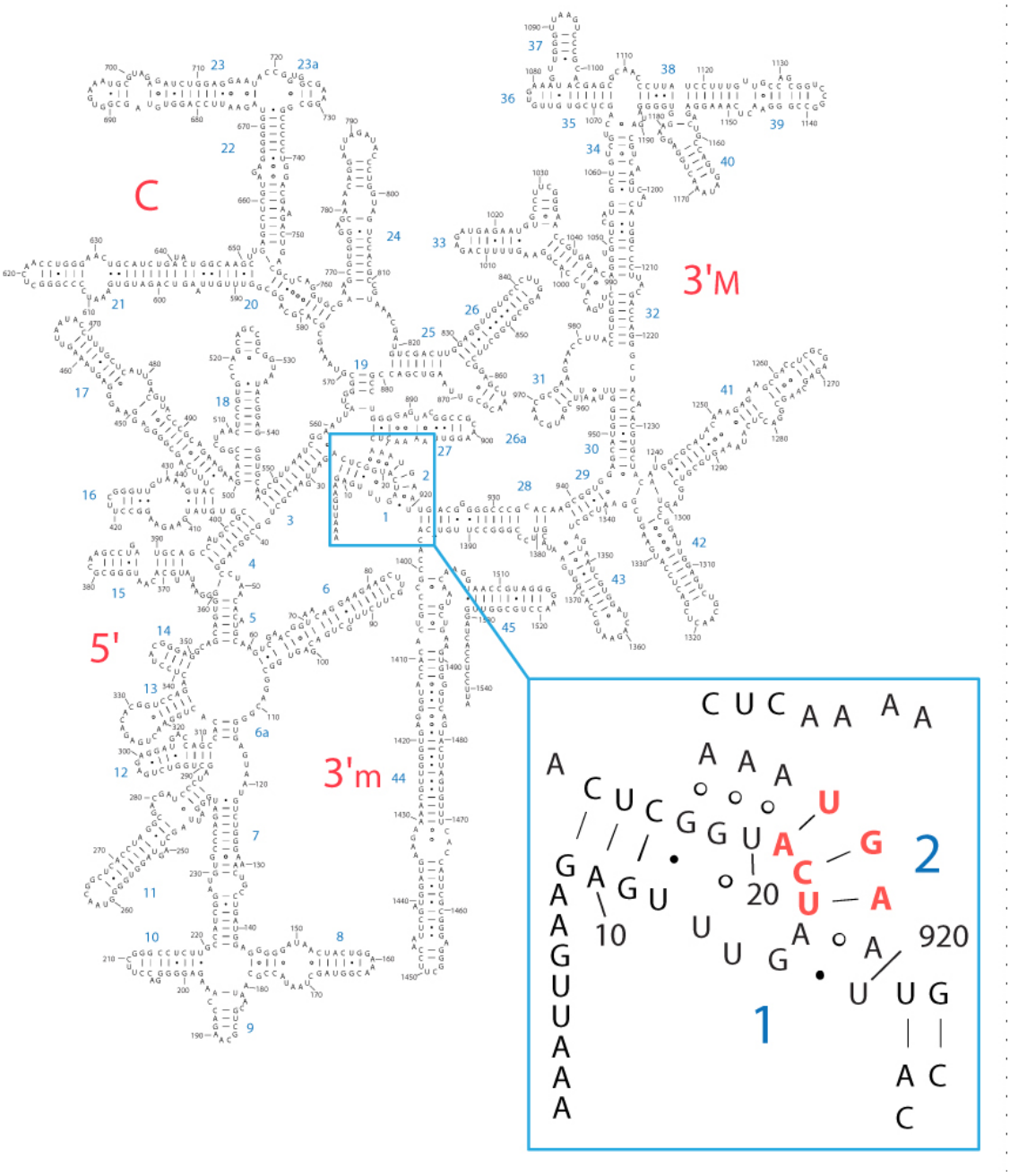

Figure 2.2 Close view of the central pseudoknot nucleotides.
Nucleotides mutated in Brink's study is shown as red. Secondaru structure is taken from in-hause Ribovision webserver.

diacetyl chloramphenicol but also labeled the protein by L-[^{35}S]methionine incorporation into CAT through specialized ribosomes in the presence of antibiotic spectinomycin. Specialized ribosomes confer resistance to

spectinomycin due to C1192U mutation in 16S rRNA while chromosomally encoded ribosomes are being blocked by the antibiotic.

After measuring the activity of CAT, Brink and coworkers suggest that the C18A, C18G, and C18U mutant ribosomes were not capable of synthesizing the specialized protein CAT. However, C18A-G917U, C18G-G917C, and C18U-G917A double mutants were showing the same activity levels as wild type ribosomes. Interestingly, single C18U mutant ribosomes show 25% activity compared to wild type ribosomes which is suggested as a formation of a weaker G-U wobble base pairing. Those results suggest that disrupting helix 2 of 16S rRNA inhibits translation.

In order to further understand the mechanism of inhibition, they performed a primer extension analysis of mutant and wild-type ribosomal rRNAs. In the presence of ddGTPs, C1192U mutation leads to a different length of cDNA product for a chromosomally encoded 16S rRNA and specialized 16S rRNA. Their results suggest that incorporation of 16S rRNA into ribosomal particles is not affected by mutation in the central pseudoknot. Moreover, by using primer extension analysis, they have tested the 5' end processing of specialized 16S rRNAs and did not found any misprocessing.

Furthermore, to understand whether a mutant 16S rRNA is capable of forming a 70S complex or polysomes, they harvested ribosomal particles from cells and separated them with sucrose gradient. After isolation of trisomes, disomes, 70S, and 30S particles, they analyzed 16S rRNA of each ribosomal unit by primer extension. Their findings suggest that 16S rRNAs containing disruptive mutations in the central pseudoknot are not incorporated into 70S or polysomes but 30S subunits (Figure 2.1). Those results elaborate, that disruptive mutations in

the central pseudoknot of 16S rRNA interferes with formation of the 70S initiation complex resulting in severely diminished or halted translational activity. However, this does not affect the 30S small ribosomal subunit formation (Brink, Verbeet, & De Boer, 1993).

2.2.1 Base complementarity of helix 2 of 16S rRNA is essential for ribosome function

With a similar effort, Poot and coworkers introduced mutations to the central pseudoknot by disrupting the first and last of the triple base pairs in helix 2 of 16S rRNA (Poot, Worm, Pleij, & van Duin, 1998). Their findings were very similar to findings of C18 mutations, located in the middle of the triple base pair. By using the specialized ribosome system as described earlier, they observed very similar effects in terms of translational activity of mutant ribosomes *in vivo*. Overall, they suggest that the central pseudoknot functions as a permanent structural unit in the center of the 16S rRNA.

2.3 The Central pseudoknot is Required for Stability of the Ribosome

To elaborate on the mechanism of inhibition caused by disruption of the central pseudoknot, Poot and coworkers used the same C18A mutation in the central pseudoknot as studied by Brink and coworkers (Poot, Pleij, & van Duin, 1996). Their findings suggested that the translational deficiency is caused by the instability of 30S subunit containing the mutant central pseudoknot.

In their report, they studied Brink and coworkers' specialized ribosome systems *in vitro*. After harvesting the cells, which governs specialized ribosomes, they observed from patterns of the sucrose gradient that C18A mutant

ribosomes are not capable of forming 70S tight couples even though mutant 30S fractions are almost as active as wild-type 30S ribosomal subunits in terms of forming a translation initiation complex. Sucrose gradient profiles indicate that the wild-type ribosomal subunits are 70S tight couple forms while mutant

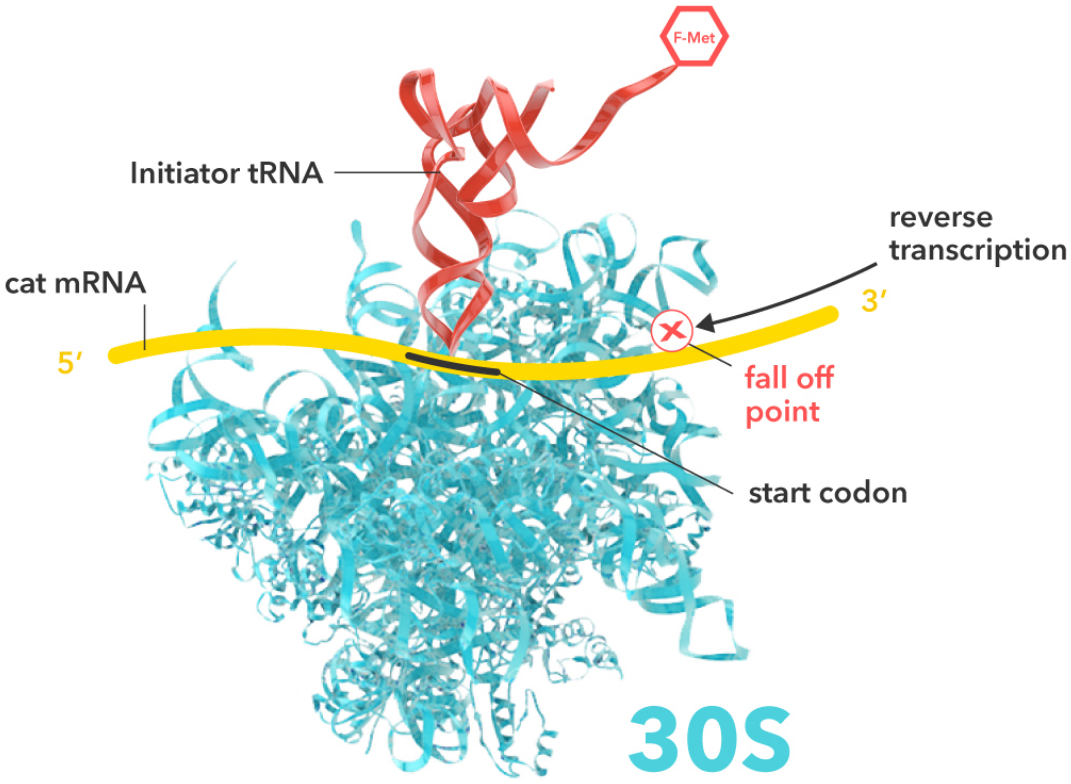

Figure 2.3 30S initiation complex representation for Poot's study.
In the presence of cat mRNA, f-met tRNA, and 30S ribosomal subuntis from 30S free fractions or 70S tight couple fractions, initiation complex formation observed by reverse transcription. Reverse transcriptase fall off 13 nucleotide downstream of the start codon indicating a toeprint. This study confirms that 30S subunits governing mutant central pseudoknot are able to form the initiation complex.

ribosomal subunits aggregate as 30S forms. By primer extension of the different fractions, they measured the population of mutant ribosomes in 70S tight couples at 78% while mutant ribosomes were at 20%.

Furthermore, they analyzed the 30S initiation complex formation for the mutant 30S subunits in presence of CAT mRNA and $tRNA_f^{Met}$ using toeprinting method. Primer extension of CAT mRNA yielded a unique product length based on the initiation complex formation. Hence, their results show that the mutant 30S subunits are capable of forming a 30S initiation complex as active as wild-type subunits. A demonstration of the toeprinting analysis is shown in Figure 2.3.

Moreover, they isolated the mutant 30S subunits by affinity chromatography. The oligonucleotides complementary to the 3' end of the chromosomal 16S rRNA are attached to a column matrix in order to separate the mutant 16S rRNA containing 30S subunits. They observed a 3 fold decrease in toeprinting efficiency after affinity purification of the mutant 30S subunits. They suggest that the loss of activity is because of the loss of ribosomal proteins. To test this hypothesis, they performed *in vitro* reconstitution of the mutant 30S subunit by incubating it with total 30S ribosomal protein extract. After reconstitution, the toeprinting activity is regenerated.

Overall, they suggested that the *in vitro* 70S complex formation is defective due to the C18A mutation in the central pseudoknot for the mutant 30S subunits. However, the mutant ribosomes are capable of forming initiation ternary complex while their stability is adversely affected by the mutation. As a result, the mutant ribosomes lose their ribosomal proteins.

2.4 Acylation of the Ribosomal Protein S5 is Affected by the Central Pseudoknot

Further contribution to understand the mechanism of translation inhibition came from Poot and coworkers in their ribosomal protein S5 study (Poot,

Jeeninga, Pleij, & van Duin, 1997). They used the same specialized ribosome system to tract proteins from C18A mutant ribosomes. Their SDS-PAGE electrophoresis of 30S ribosomal proteins from fractions containing wild-type ribosomes and C18A mutant ribosomes from 70S tight couples, C18A mutant ribosomes from 30S free subunits, and U1065C-A1191G mutant ribosomes from 30S free subunits reveals that a deficiency of ribosomal proteins S1and S2 occurs in mutant ribosomal subunits. However, this deficiency is not only correlated with C18A mutation but also U1065C and A1191G double mutation in the upper stem of helix 34. In order to distinguish the effect of C18A mutation in the central pseudoknot, they performed 2D gel electrophoresis, where the mobility of proteins were depended on their charge.

A deficiency of ribosomal proteins S18 and S21 was observed in the C18A mutant ribosomes as well as in U1065C and A1191G double mutants, indicating that deficiencies of ribosomal proteins S18 and S21 are not related to C18A mutation. However, they observed a change in pattern for the ribosomal protein S5, which is unique to C18A mutant ribosomes derived from 30S free subunits. A spot in the 2D electrophoresis gel from C18A mutant ribosome corresponding to the ribosomal protein S5 is split into 2 parts. This separation pattern was not observed in SDS-PAGE analysis. They suggested that the split spot for the ribosomal protein S5 is caused by a change in net charge of the protein rather than by a significant change in molecular mass.

In order to characterize the unknown pattern of ribosomal protein S5, they performed Western blot analysis by using antibodies against S5. They separated ribosomal proteins from different fractions in lower percentage acrylamide gel.

The results suggested that unknown pattern and also known pattern of S5 reacted specifically with the S5 antibodies. The mobility of S5 derivative was slightly higher than the mobility of S5. Interestingly, S5 derivative is observed in all fractions except the wild-type central pseudoknot. However, the ratio between 70S derived C18A mutant ribosome and 30S derived C18A mutant ribosome suggested that 70S tight couple hinders the formation of S5 derivative compared to 30S free subunits. Based on their results, they suggested that C18A mutation in the central pseudoknot in free 30S subunits causes an overall increased positive charge of ribosomal protein S5. Since the ribosomal protein S5 carries an acylated N-terminal alanine residue, absence of acylation would cause an increase in positive charge. They further suggested that mutation in the central pseudoknot might change the S5 binding site of the 16S rRNA without affecting the stoichiometry. Overall, they concluded that a modification of the ribosomal protein S5 takes place in the ribosome and mutations in the central pseudoknot inhibit this modification.

2.5 A Conformational Switch in Helix 27 of 16S rRNA

Another important hypothesis about the central pseudoknot came from Lodmell and Dahlberg who proposed an alternative conformation of helix 27 during decoding process (Lodmell & Dahlberg, 1997). They hypothesized that three closing stem base pairs of helix 27 switch to the adjacent Watson-Crick counterparts during decoding of messenger rRNA.

In the intact SSU, nucleotides C912, U911, and C910 pair with G885, G886, and G887 respectively. The alternative base pairing pattern suggests that nucleotides C912, U911, and C910 pair with G888, A889, and G890 respectively

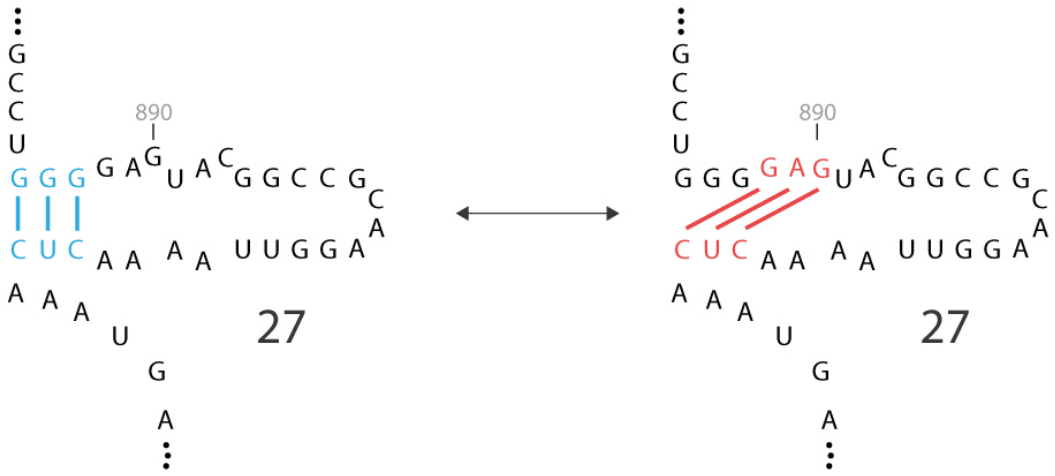

Figure 2.4 Conformational switch in the stem edge of helix 27 of SSU proposed by Dahlberg and Lodmell.
Native 912-885 conformation is shown on the left with symbolic base pair interactions (blue). Alternative 912-888 conformation is shown on the right with symbolic base pair interactions (red).

(Figure 2.4). In order to experimentally test their hypothesis, they made side directed mutants of those nucleotides to observe an equilibrium shift between the two conformations. They distinguished those mutations as restrictive mutants, which favor the 912-888 conformation, and error-prone mutants, which favor 912-885 native conformations. Furthermore, they assayed translational fidelity by measuring the rate of aberrant read-thorough in-frame stop codons in the reporter gene lacZ. The results suggest that 912-885 native conformation mutants had higher rates of reading through in-frame stop codon while 912-885 alternative conformation mutants had lower rates. However, those alternative conformation mutants had an increased rate of frameshifting.

Furthermore, their rRNA mutation results were similar to the mutation experiments with ribosomal proteins. Specifically, mutations in the ribosomal protein S12, which confer resistance to error-inducing antibiotic streptomycin,

retard translation rate and increase accuracy. Mutations in ribosomal protein S5 increase the basal translational error rate. From those results, they suggest that 912-888 favoring mutations are compatible with S5 mutations while 912-885 favoring mutations are compatible with S12 mutations. They then hypothesize that reciprocity between ribosomal RNA and ribosomal protein mutations support the idea that those two confirmations are not just static but dynamic.

Moreover, their chemical modification probing experiments provided physical evidence to their hypothesis. Kethoxal and dimethyl sulfate modifications in the RNA specifically alter guanines and adenines, which do not involve in base pairing. Probing the changes in reactivity revealed that both confirmations are supported by the data they obtained from chemical modifications.

2.6 Arrangements of the Central pseudoknot Region of 16S rRNA

In the absence of high-resolution crystal structures, Juzumiene and Wollenzien studied the central pseudoknot of 16S rRNA with 4-thiouridine crosslinking to understand its interactions with other regions of the 16S rRNA (Juzumiene & Wollenzien, 2001). They substituted uracil bases in the first 20 nucleotides of 16S rRNA with 4-thiouridine (s^4U). They synthesized the 1-20 nucleotide fragment of 16S rRNA *in vitro* and incorporated s^4U into uridine positions 5, 14, 17, and 20 one at a time by chemical synthesis. The 21-1542 containing fragment of 16S rRNA was synthesized *in vitro* and ligated to the 1-20 fragment. Reconstitution of 30S subunit was made by incubating modified 16S rRNA with 30S ribosomal proteins. Near-UV irradiation produced crosslinking in 30S subunit containing s^4U in the central pseudoknot of 16S rRNA. After crosslinking,

they analyzed the sites with reverse transcription and separated the fragments with gel electrophoresis. Their results suggested that the central pseudoknot is in close contact with helix 27, decoding center, and a part of domain III RNA.

After high-resolution crystal structures emerged, the structure of the entire ribosome was resolved and those results became visually available.

2.7 A Functional Relationship Between Helix 1 of the Central Pseudoknot and Helix 27 Tetraloop

Belanger and coworkers studied the central pseudoknot of 16S rRNA with several mutations using genetic complementation approach (Belanger, Theberge-Julien, Cunningham, & Brakier-Gingras, 2005). In their previous report, they showed that the mutation in the loop of helix 27 affects translation *in vivo* (A900G). Since the loop of helix 27 is one of the intersubunit bridges, which connect SSU and LSU via A-minor interactions, mutations at A900 will affect intersubunit assembly and ultimately efficiency of translation.

In their genetic complementation approach, they introduced random mutations to 16S rRNA alongside A900G and observed survival bacterial colonies. Their strategy to introduce random mutations was either using mutagenic XL1-Red *E. coli* strain or error-prone PCR. As described earlier in Poot and Brink's studies, Belanger and coworkers used similar specialized ribosome system techniques to monitor only mutant translation systems without disturbing *E. coli*'s native translation. From their findings, they observed only two specific mutations, which assisted A900G mutant ribosomes to gain function again. Those mutations were U12C or deletion of U12. This nucleotide position is the corresponding helix 1 of the central pseudoknot (Figure 2.5).

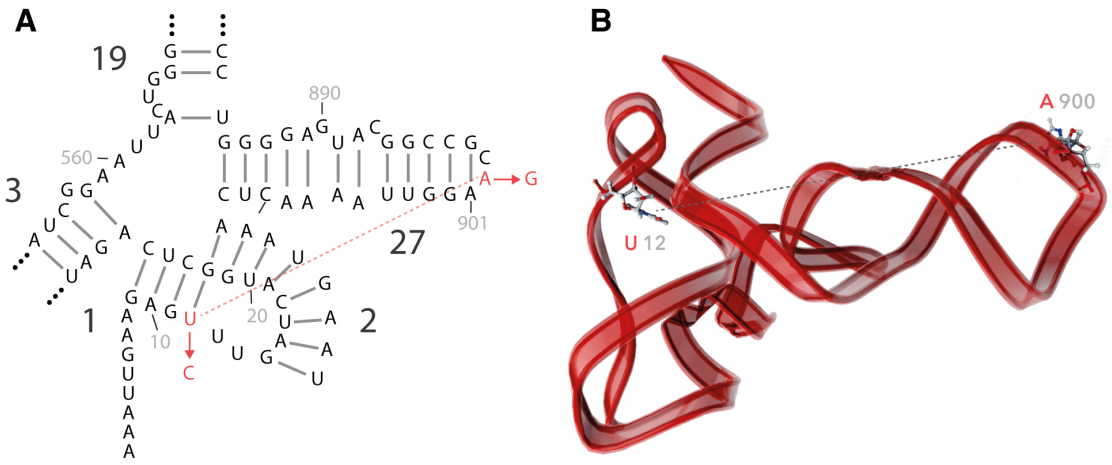

Figure 2.5 Functional relationship between helix 1 and helix 27 tetraloop. A) Secondary structure map indicates the location of the compensating mutations in helix 1 and helix 27 (demonstrated as red). B) Three-dimensional structure shows the positions in 3D space. The structure is represented as ribbons while nucleotide postions A900 and U12 are represented as sticks. Three-dimensional structure is extracted from 16S rRNA (PDB: 2J00).

According to their results, mutation in helix 1 affect the helix 27 tetraloop indirectly. Their hypothesis is that the mutations in helix 1 destabilize helix 2, which ultimately makes the central pseudoknot more flexible. Increased flexibility of the central pseudoknot leads to increased flexibility of helix 27 which could have less hindrance in its orientation in terms of intersubunit assembly. However, we know from previous research that disrupting helix 2 causes inhibition of proper ribosome functioning. This data also supports the idea that the central pseudoknot is vital for ribosome function.

2.8 Conclusions

Proteins are synthesized by ribosomes after a highly complex series of molecular events. SSU plays a major role in protein synthesis by decoding the mRNA. During translation, SSU rRNA adopts different conformations to provide the best molecular atmosphere in terms of subunit association, decoding, and

translocation. The central pseudoknot in the central core of the SSU rRNA is involved in those processes structurally and functionally. In this chapter, we have summarized research that was conducted on the central pseudoknot.

Mutational experiments on the central pseudoknot show that an ablation of the central pseudoknot inhibits the whole translation process. The mutant SSU rRNA incorporates into 30S particles, but it cannot interact with the 50S particle to form tight 70S complexes. The inhibition of translation suggests a defect in subunit association. However, 30S subunits containing the mutant central pseudoknot can form f-met tRNA and mRNA initiation complex. Furthermore, acylation of ribosomal protein S5 is affected by the mutations in the central pseudoknot. Moreover, a conformational switch near the central pseudoknot is suggested during translation, which indicates the flexibility of the central pseudoknot. Functional relationship between the helix 1 of the central pseudoknot and ribosomal intersubunit bridge B2c also stresses the significance of the central pseudoknot in protein synthesis.

2.8.1 The contribution of this thesis on the small ribosomal subunit and the central pseudoknot

In the following chapters of this thesis, we will discuss my research and contribution on the small ribosomal subunit and more specifically the core region of the SSU (as well as the central pseudoknot).

After 30 years of research on the central core of the SSU, we do not have a coherent answer for the role of the central core of the SSU in the ribosome function and architecture. There were still a lot of mistakes in structures and those hinder our understanding of the structure and function of the ribosome. After

high-resolution crystal structures were solved for the ribosome, molecular interactions become easy to see. By this connection, first of all we have revised the secondary structure maps of rRNAs and defined a RNA domain. Domain criteria for the SSU reveals a central core domain which all the other historical domains radiate from it. Those results help better understanding of the SSU architecture and function.

CHAPTER 3

REVISION OF THE SECONDARY STRUCTURE MAP OF THE SSU RNA

Secondary structure maps are powerful tools to simplify the structures of large RNA molecules. Three-dimensional structures can be complicated to comprehend. Therefore, combining 3D structures and secondary structure maps provides deeper and faster understanding of large RNA molecules together with additional information such as base pairs and their types, and tertiary interactions.

3.1 Introduction

A secondary structure map of RNA is a sequential arrangement of primary structure by a prediction or an observation. While RNA molecules fold into complex motifs, secondary structure maps help to reflect those secondary and tertiary motifs in two dimensions.

RNA secondary structures symbolize a variety of information such as base pairs, double helices, loops, bulges, and single-strands. Specifically, 2D structures play an important role in understanding ribosomes, which are extremely large and highly complex molecules.

3.1.1 Historical secondary structure maps of ribosomal RNAs

Predicting RNA secondary structures was a highly challenging task before 3D structures emerged. Woese, Gutell, Noller and their coworkers proposed the first secondary structures for the 16S rRNA and later for 23S rRNA in early 1980s (Noller et al., 1981; Noller & Woese, 1981; Woese et al., 1980b). Comparative

sequence analysis supported by chemical and enzymatic probing provided information to predict those highly complex RNA molecules.

Co-variation analysis uses a rich sequence database as primary input and provides powerful and widely applicable determination of rRNA secondary structures when 3D information is not present. Co-variation analysis produces accurate Watson-Crick base pairs and very few false-positive base pairs (Gutell, Lee, & Cannone, 2002). However, structure determination by co-variation analysis has limitations due to a lack of 3D information. Co-variation methods have difficulties to accurately reveal non-canonical base pairs, specifically purine-purine base pairs. For instance, Helix 26a of 23S rRNAs was left out in the historical secondary structure map of LSU rRNA predicted by co-variation analysis (Cannone et al., 2002; Gutell et al., 2002). Helix 26a region of 23S rRNA is represented as extended single stranded RNA instead of a helix in traditional secondary structure maps. In contrast, helix 26a is universally conserved and thermodynamically stable and plays an important structural role in 23S rRNA (Leontis & Westhof, 1998; Serra et al., 2002). The domain architecture of the 23S rRNA was misrepresented due to the missing helix 26a and recently revised and corrected by Petrov and coworkers (Petrov, et al., 2013).

3.1.2 Relationship between 2D and 3D structures

After high-resolution 3D structures of ribosomal subunits were revealed, base pair interactions within ribosomal RNA nucleotides became easy to investigate. Observations of the 3D structures show coaxial helical stacks and non-canonical base pairs, which were missing in co-variation analysis. With an analogous effort to the previous 23S rRNA secondary structure revision (Petrov et

al., 2013), we have re-determined the secondary structure of 16S rRNA by using a three-dimensional approach (Petrov et al., 2014).

Our focus is to accurately re-define the secondary structure of the SSU rRNAs. Co-variation based traditional secondary structures were widely used and accepted. For this reason, we modified the traditional *E. coli* 16S rRNA secondary structure and incorporated non-canonical base pairs. The most significant change is in the central pseudoknot of the SSU rRNA. In historical secondary structures, the central pseudoknot is represented as three Watson-Crick base pairs. However, there are several non-canonical base pairs in the central pseudoknot as observed by X-ray crystal structures (Selmer et al., 2006). Co-variation approaches are especially problematic for highly idiosyncratic RNA sequence regions such as expansion segments, because appropriate sets of alignable sequences may not be available or readily identifiable (Petrov et al., 2014).

3.2 Materials and Methods

Atomic coordinates were obtained from Protein Data Bank (PDB). Base-pairing and base-stacking interactions were obtained from the library of RNA interactions (FR3D) (Sarver, Zirbel, & Stombaugh, 2008) and confirmed by inspection and in-house Matlab script. The co-variation *E. coli* secondary structures of LSU and SSU rRNAs were obtained from http://rna.ucsc.edu/rnacenter/ribosome_images.html, adjusted and extended using the program XRNA (http://rna.ucsc.edu/rnacenter/xrna/xrna.html), and finalized with Adobe Illustrator, written out as svg and png files. The secondary structures of all other species were built from the *E. coli* template. We use historical

representations as much possible, exceptions are where conflicts arise with correct helical assignments or strand continuity.

E. coli secondary structures were determined from X-ray structures of Cate and coworkers (Dunkle et al., 2011a) (PDB entries 3R8S, 4GD1). *T. thermophilus* secondary structures were determined from X-ray structures of Ramakrishnan and coworkers (Selmer et al., 2006) (PDB entries 2J00, 2J01). *S. cerevisiae* secondary structures were determined from X-ray structures of Yusupov and coworkers (Ben-Shem, Garreau de Loubresse, et al., 2011)(PDB entries 3U5B, 3U5C, 3U5D, 3U5E). *D. melanogaster* and *H. sapiens* secondary structures were determined from cryo-EM structures of Beckmann and coworkers (Anger et al., 2013) (PDB entries (3J38, 3J3C, 3J39, 3J3E for *D. melanogaster*; PDB entries 3J3A, 3J3B, 3J3D, 3J3F for *H. sapiens*).

3.3 Results and Discussion

Ribosomal RNA secondary structures previously determined by a variety of methods including co-variation analysis (Fox & Woese, 1975; Noller et al, 1981; Woese et al., 1980a), chemical probing (Deigan, Li, Mathews, & Weeks, 2009; Hajdin et al., 2013; Siegfried, Busan, Rice, Nelson, & Weeks, 2014; Weeks, 2010), thermodynamic predictions (Zuker, 2003) and by geometric analysis of molecular interactions within 3D structures (Petrov et al., 2013). We have developed a series of revised rRNA secondary structures from 3D structures, by improving clarity, accuracy, and usability. The major drawback in structural approach is a limited number of high-resolution 3D ribosome structures. However, advances in structural biology and biochemistry bring an increasing number of ribosome 3D structures with it (Anger et al., 2013; Armache et al., 2010; Ben-Shem, Jenner,

Yusupova, & Yusupov, 2010). Recent available ribosome structures from all kingdoms of life render the structure determination by geometric analysis as a strong tool for rRNA secondary structures.

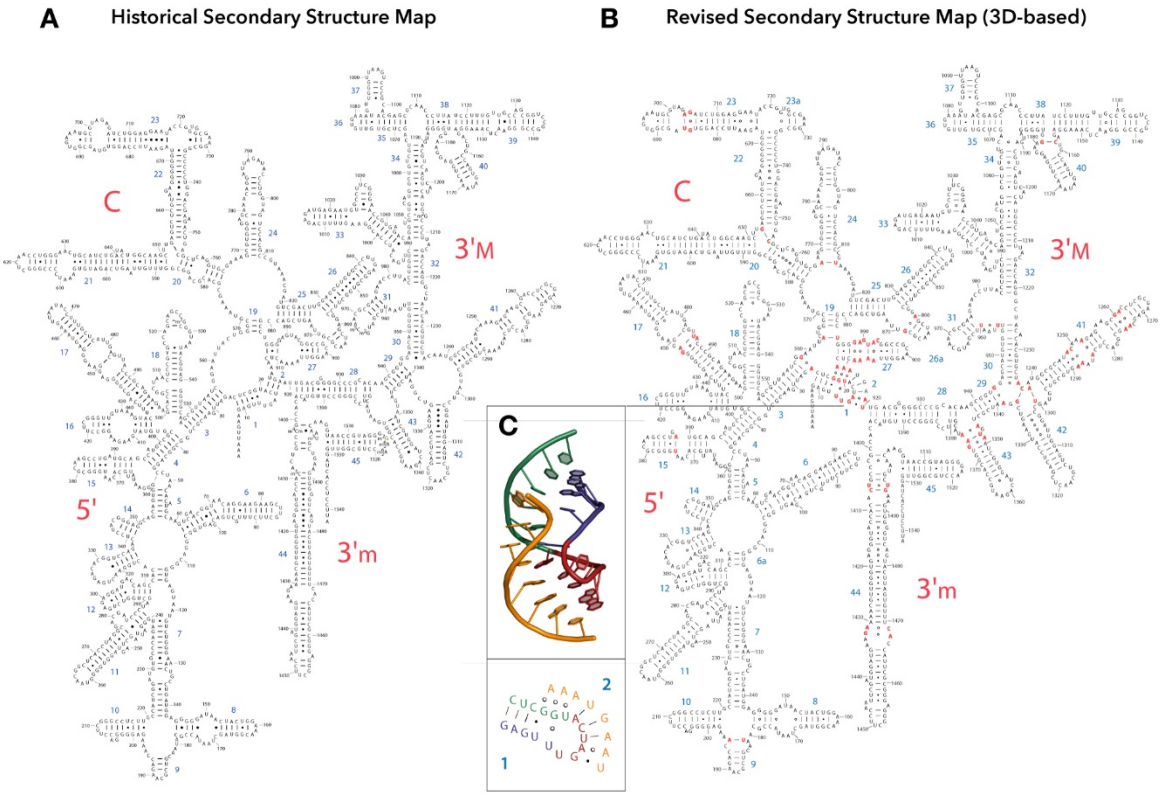

Figure 3.1 Revised and the historical secondary structure of 16S rRNA. (A) Historical secondary structure of *E.coli* 16S rRNA. Helix (purple) and nucleotide numbers (black) and domains are indicated (red). (B) Revised 3D-based secondary structure of *E.coli* 16S rRNA. Helix (purple) and nucleotide numbers (black) and domains are indicated (red). Red colored nucleotides represent the changes made over the historical secondary structure. (C) Closer view of the helix 1 and 2 in 3D and 2D.

RNA helices are the major defining units of RNA secondary structures (Butcher & Pyle, 2011; Richards, 1969). We identify helices by specific criteria previously defined by Petrov and coworkers (Petrov et al., 2013). According to geometric and molecular interactions criteria, a RNA base can be in two

discrete states: paired or non-paired (Leontis, Stombaugh, & Westhof, 2002; Macke et al., 2001). A paired base can involve in motifs such as secondary interactions, tertiary interactions, or both. As described by Levitt (Sim & Levitt, 2011), we define helices as base-paired nucleotide series bound by non pairing nucleotides. Furthermore, in some cases, decision of which nucleotide belongs to which helix was made by base stacking criteria. Incorporating stacking information was made by observation of 3D structures, and therefore we define helices as base pairs in the form of a continuous base-paired stack that carries its connectivity. In our definition of a helix, bulges or non-pairing defects do not break the helix rule unless they change stacking and connectivity. Each nucleotide belongs to one unique helix. Non-canonical and canonical base pairs are incorporated together as 3D information indicates.

Simple helical definition of secondary structures (Richards, 1969) separate nested and non-nested helices (Rivas & Eddy, 2000; Searls, 1992; Waterman & Smith, 1978). Helices between eukaryotic expansion segments (as in 18S rRNAs of *S. cerevisiae, D. melanogaster*, and *H. sapiens*) are the longest non-nested helices. Generally, non-nested helices (kissing loops and pseudoknots) are categorized as tertiary interactions (Butcher & Pyle, 2011; Smit, Rother, Heringa, & Knight, 2008).

In our 3D -based secondary structures, secondary and tertiary helices are defined as the nest/non-nest definition. Furthermore, thermodynamic stability of folded RNA conferred by all pairing interactions is the major criteria defining RNA secondary structure. To this end, in our approach we incorporated the non-canonical and canonical base pairs together as a result of helix definition.

3.3.1 Looking at the central pseudoknot

When it comes to the central pseudoknot of the 16S rRNA (Pleij, Rietveld, & Bosch, 1985), we define it as helices 1 and 2. Following the original representation of Woese (Woese et al., 1980a), we define helix 2 as a secondary element despite the fact that it is non-nested. The central pseudoknot is pyhlogenetically and structurally highly conserved (Gutell, Larsen, & Woese, 1994) and is an essential component of SSU RNA. Formation of the central pseudoknot is directly related with SSU stability and ribosome function, and represents an irreversible step of SSU maturation (Segerstolpe et al., 2013).

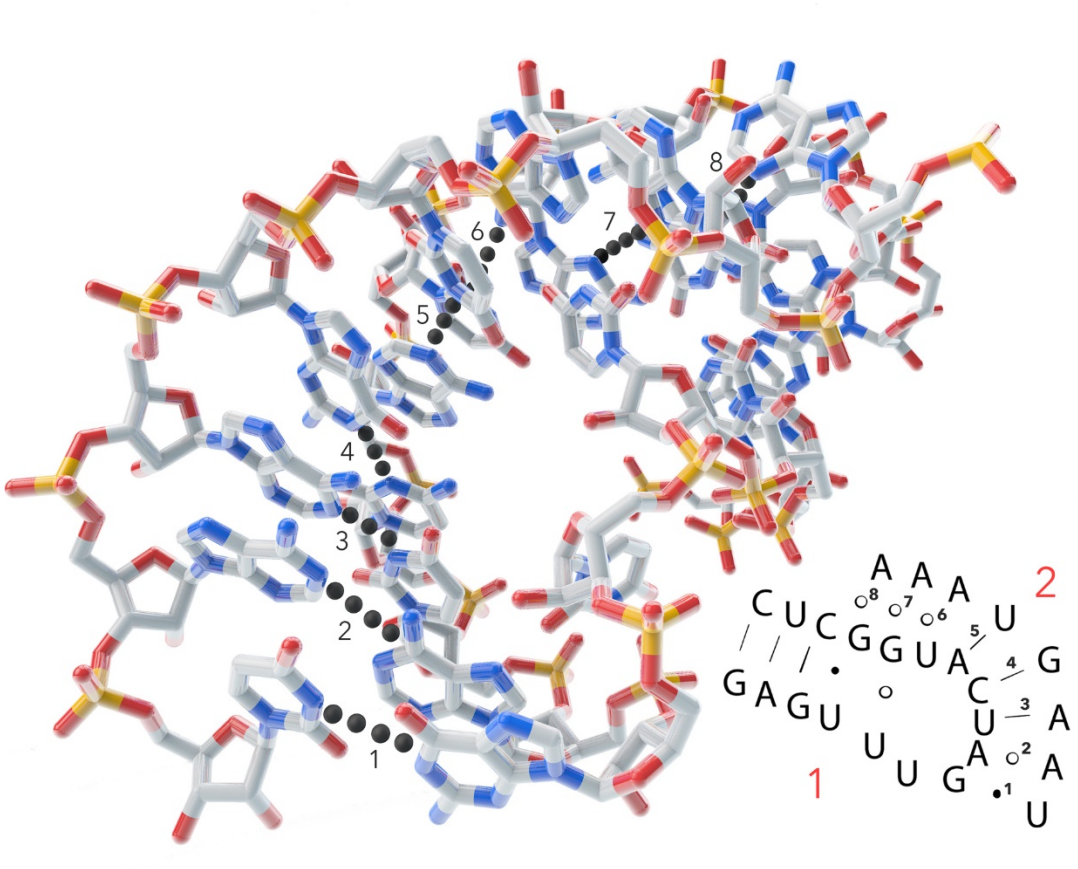

Figure 3.2 A closer look at the central pseudoknot.
The canonical and non-canonical base pairing interactions in helix 2 of the central pseudoknot is shown as black dots and numbered to match the secondary structure representation. Structure is from E.coli 16S rRNA. (PDB ID:4GD1)

The co-variation secondary structure of the central pseudoknot is missing vital non-canonical and also canonical base pairs. Figure 3.2 illustrates the entire base pair interactions in helix 2. Our main revision over traditional secondary structure maps is the central pseudoknot region and modifying its base-pairing interactions as revealed by 3D structures. The central pseudoknot contains conserved triplets of bases in the nucleotides U12-G22-A912 and U13-U20-A914. In our revised secondary structure, these base triples are represented as double base pairs in each side (Figure 3.1 B and C). This representation allows interpreting the pseudoknot in more detail and enhances clarity in recognizing all base pairing interactions in the central pseudeoknot. The representation used here was formulated by Brakier-Gingras and coworkers (Belanger et al., 2005) and by Gregory and Dahlberg (Gregory & Dahlberg, 2009) using information from 3D crystal structures. Additionally, Westhof and Lescoute correctly drew the central pseudoknot in their information-rich wiring diagrams (Lescoute & Westhof, 2006). Gutell recently revised the historical secondary structure of 16S rRNA to adjust the central pseudoknot and incorporate many of the non-canonical base pairs (Weijia et al., 2011). The central pseudoknot's revised representation can be incorporated into the historical secondary structure maps without major rearrangements necessary. The 3D structure based revised secondary structure of the 16S rRNA of *E. coli* is shown in Figure 3.1 together with the historical secondary structure. The red colored nucleotides in the 3D based secondary structure indicate changes over the historical secondary structure.

3.4 Conclusion

We have revised secondary structures for the 16S/18S rRNAs of *E. coli, T. thermophilus, S. cerevisiae*, *D. melanogaster*, and *H. sapiens* by using 3D-structure based approach. We made secondary structures mapped with a variety of data such as base pairing, domains and base stacking available to researchers through our server: http://*apollo.chemistry.gatech.edu/RibosomeGallery*. With this available improvement, researchers are able to process their data faster and in more detail, allowing them to interpret their findings more accurately.

CHAPTER 4

REVISION OF THE DOMAIN STRUCTURE OF SSU RNA

We propose and validate a new architectural model of the ribosomal small subunit, with broad implications for function, biogenesis and evolution. We define an rRNA domain: compact and modular, stabilized by self-consistent molecular interactions, with ability to fold autonomously. Each rRNA helix must be allocated uniquely to a single domain. These criteria identify a core domain of small subunit rRNA (domain A), which acts as a hub, linking to all other domains by A-form helical spokes. Domain A, which exhibits elements of tRNA mimicry, is the essential core of the small ribosomal subunit. Understanding the structure and dynamics of domain A will provide valuable insight into the translational machinery.

4.1 Introduction

The ribosome is a ribonucleoprotein complex that conducts one of life's most ancient and universal processes, synthesizing proteins. The large ribosomal subunit (LSU) contains the peptidyl transferase center (PTC) and catalyzes transpeptidation. The small ribosomal subunit (SSU) contains the decoding center and reads the messenger RNA (mRNA). Much of ribosomal function is performed by ribosomal RNAs (rRNAs) (Ban, Nissen, Hansen, Moore, & Steitz, 2000; Noller, Hoffarth, & Zimniak, 1992) while the ribosomal proteins act primarily as structural stabilizers (Ramakrishnan & White, 1998). Our understanding of translation has advanced over the last decade and a half with the explosion in sequences and

by the determination of three-dimensional structures (Ban et al., 2000; Cate, Yusupov, Yusupova, Earnest, & Noller, 1999; Harms et al., 2001; Selmer et al., 2006). X-ray crystallography and cryo-electron microscopy (Cryo-EM) have provided atomic resolution structures of ribosomes from all three domains of life (Anger et al., 2013; Ban et al., 2000; Frank et al., 1995; Selmer et al., 2006). In prokaryotes, the SSU is composed of the 16S rRNA and 20 ribosomal proteins (rProteins) (Carter et al., 2000). The secondary structure of SSU rRNA, determined initially from co-variation (Noller & Woese, 1981), is conventionally shown with three or four domains, directly linked to each other at a common origin (Frank et al., 1995; Petrov, et al., 2014; Woese et al., 1980b).

RNA secondary structures, with symbolic representations of base pairs, double-helices, loops, bulges and single-strands, are frameworks for understanding structure, folding and function, and for organizing a wide variety of information. Secondary structures reveal how rRNA is organized into quasi-independent domains, which can be used to infer mechanisms of assembly and evolution. Rigorous inference of secondary structure and domain organization can now be accomplished using high-resolution three-dimensional structures. Petrov and coworkers have previously described criteria for defining an rRNA domain in three-dimensional structures (Petrov et al., 2013). The rRNA comprising a domain must be compact and modular, with a self-contained and integrated system of stabilizing molecular interactions. A given rRNA helix must be allocated uniquely to a single domain. A domain must fold autonomously when excised from the surrounding rRNA.

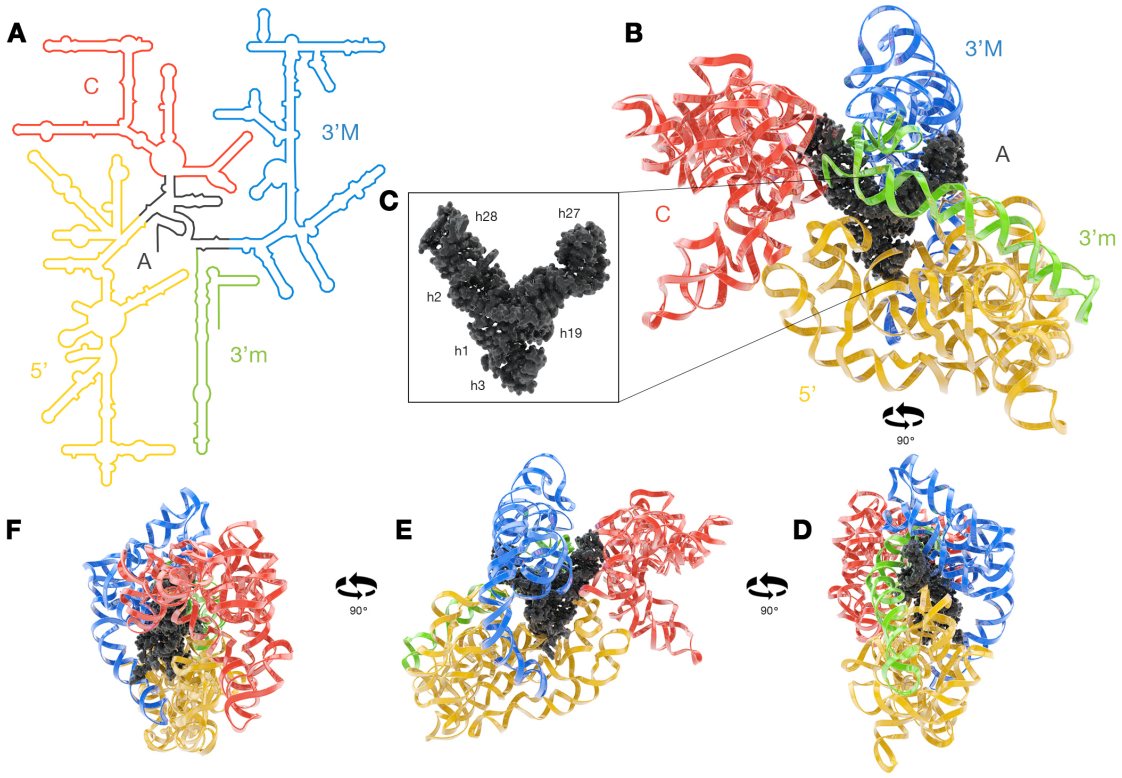

Figure 4.1 Revised domain structure of SSU rRNA from *T. thermophilus*.
(A) Secondary structure of the *T.thermophilus* 16S rRNA. The domains are colored. Domain A is black, the 5' domain is yellow, the central domain is red, the 3' major domain is blue, and the 3' minor domain is green. (B) Three-dimensional structure of the SSU rRNA (PDB ID 2J00) in a series of 90° rotations. The rRNA is represented in ribbon, except for domain A, which is in space filling representation. The domains in the three-dimensional representation are colored by same scheme as in the secondary structure. (C) Space filling representation of domain A. Helix numbers are indicated. (D, E, F) The figures demontrate 90^0 rotations on y-axis.

In addition we have inspected the three-dimensional structure of domain A and see analogies with other biological RNAs.

4.2 Material and Methods

4.2.1 Superimposition

PDB IDs: 2J00, 4GD1, 3U5B, 3J3C, 3J3D for *Thermus thermophilus, Escherichia coli, Saccharomyces cerevisiae, Drosophila melanogaster, Homo*

sapiens were obtained from Protein Data Bank (Anger et al., 2013; Ben-Shem, Loubresse, et al., 2011; Dunkle et al., 2011a; Selmer et al., 2006). Structures were superimposed pairwise using PyMol "super" command with default settings.

4.2.2 Shannon entropy

A fraction of nucleotide type i (C, G, A or U) was calculated by a multiple sequence alignment of the 16S/18S rRNAs. Fractional occupancy of a nucleotide in the aligned sequences gives the probability (p_i) of a nucleotide type at a given position. The Shannon Entropy (H) was calculated from the probabilities p_i of each position according to the equation below (Reza, 1994; Shannon, 2001).

$$H = -\sum_{i=1}^{4} p_i \log_2 p_i$$

The Shannon Entropy changes from 0 to 2. The lowest value means that the nucleotide type at a given position is not changing through the species, defining the nucleotide as universally conserved. The highest value indicates the equivalent population of the nucleotide type at given position.

4.2.3 Domain criteria

The domain criteria were defined by self-consistent molecular interactions such as base-pairing, base-stacking, base-phosphate and base-sugar interactions.

The domain boundaries were defined to minimize interactions between domains and to maximize interactions within domains. Each rRNA helix must be allocated uniquely to a single domain and each nucleotide must be allocated

uniquely to a single helix. A domain should have the ability to fold in native state independent from surrounding RNA.

4.3 Motivations for Research

4.3.1 The historical domain structure does not represent the 3D structure

In the historical domain structure of SSU RNA, the 16S rRNA was first divided into 3 domains, 5' domain, central domain, and 3' domain. Later, 3' domain was divided to 3' major and 3' minor domains. Those domain definitions are problematic in that the structure of the 16S rRNA is not correctly represented in the historical domain definition. For instance, helix 2 of the central pseudoknot belongs to two different domains with a separation of its strands. Helix 2 is structurally and phylogenetically highly conserved and crucial for small subunit assembly (Brink et al., 1993). Dividing such an important helix into different domains causes confusions in terms of understanding the architecture of the SSU.

Moreover, helix 2 and helix 28 have a stacking continuity in the core of the SSU. In the historical domain representation, helix 28 belongs to a different domain resulting a misrepresentation of the core structure of SSU (Figure 4.2). Furthermore, previously employed domain boundaries of each domain were not clearly stated and defined. The various different drawings for the domain boundaries were all not consistent with continuity and integrity of the central pseudoknot region of the 16S/18S rRNA. To this end, we have revised the domain structure and defined the criteria for domain definition of SSU RNA.

4.4 Results and Discussion

4.4.1 Introducing domain A: Small ribosomal subunit domains radiate like helical spokes from domain A

We have re-determined SSU domain architecture based on secondary and three-dimensional structures, providing a coherent scheme for

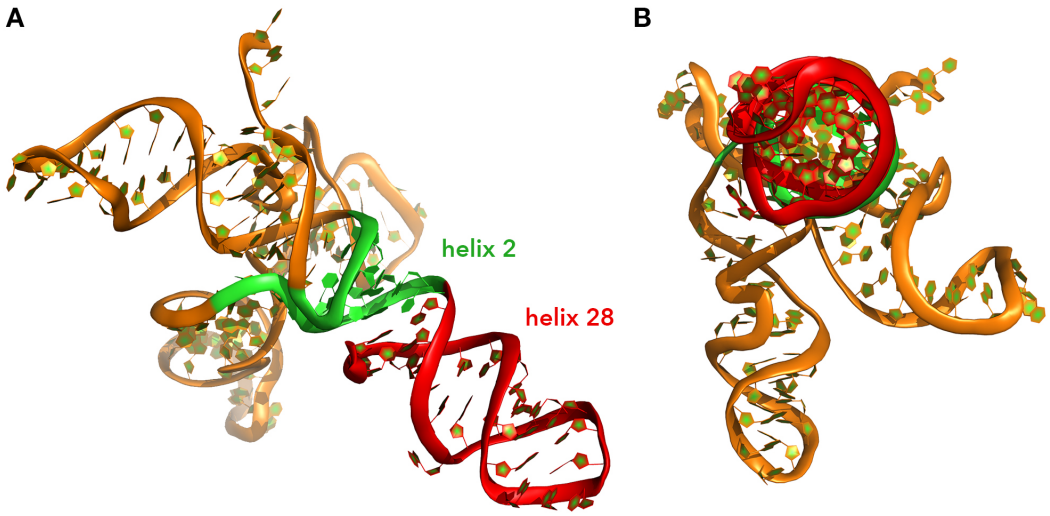

Figure 4.2 Stacking interactions between helix 2 and helix 28 of domain A^{ISO} (and 16S rRNA).
(A) Three dimensional model and cartoon representation of domain A^{ISO} rRNA indicating continuous stack between helix 2 and helix 28. Helix 2 is shown green and helix 28 is shown red while the rest of the molecule is orange. (B) Cartoon representation of the same rotated model to demonstrate continuous stack through the helical axis.

understanding SSU function. We propose an SSU domain structure in which an organizational hub is formed by domain A (Figure 4.1). Domain A links to the remnants of historical domains (the central domain, 3'M domain, 3'm domain, and the 5' domain). Each of these peripheral domains connects to domain A by a spoke, rather than directly with each other at a common center, as in previous renditions. Helix 3 is a spoke linking domain A to the 5' domain. Helix 19 is a spoke

linking the central domain while helix 28 is the spoke linking the 3' major domain. The 3' end of domain A is a spoke linking 3' minor domain (Figure 4.1). Our domain model explains the dynamical properties of the SSU. The spokes are relatively flexible, allowing the domains to move relative to each other during initiation and translocation (Valle et al., 2003). Domain A incorporates the central pseudoknot and consists of helices 1, 2, 3, 19, 27, and 28.

4.4.2 Conservation of Domain A

The structure of domain A is conserved in all ribosomes. We superimposed SSU rRNAs from bacterial and eukaryotic domains of life, including *T. thermophilus, E. coli, S. cerevisiae, D. melanogaster,* and *H. sapiens* (Figure 4.3) (Anger et al., 2013; Ben-Shem, Loubresse, et al., 2011; Dunkle et al., 2011a; Selmer et al., 2006). The root-mean square deviation (RMSD) of backbone atoms of domain A in this superimposition is only 0.78 Å), consistent with a high degree of conservation of conformation (Table 4.1). The greatest deviations are seen in the 5' terminal region, which is single-stranded (Figure 4.3). In addition, we have aligned sequences from 134 species from all three domains of life, and have calculated mutational Shannon entropies. For most of domain A, the sequences are universally conserved, with very low Shannon entropies. The sequences are most divergent in helix 3 and the 5' single stranded end (Figure 4.4).

4.4.3 Domain A shows tRNA mimicry

Domain A exhibits structural similarities with tRNA. Certain elongation factors, viral RNAs and bacterial non-coding RNAs mimic tRNAs in various ways

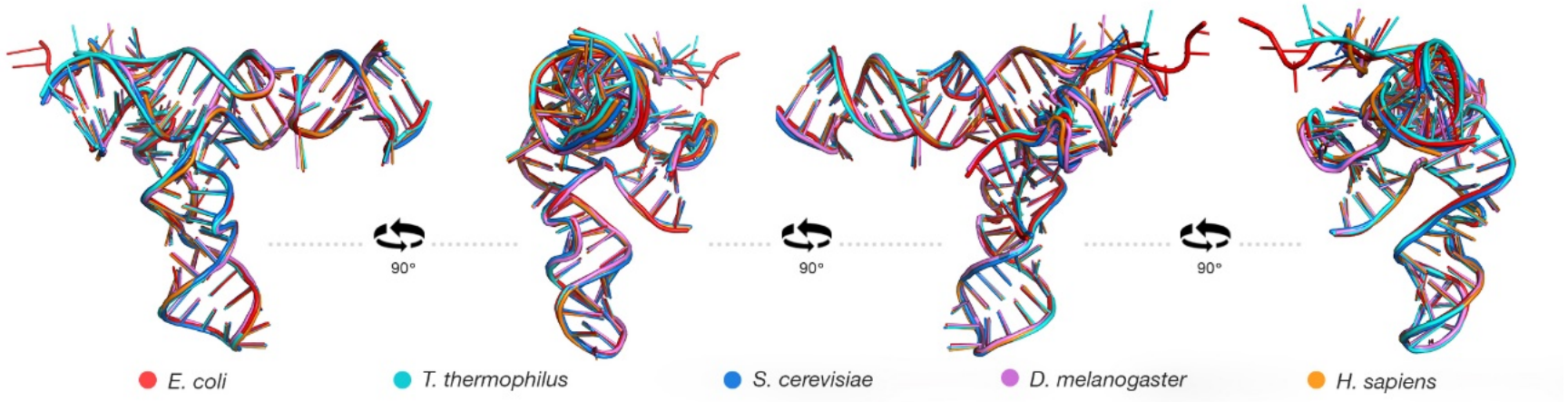

Figure 4.3 Conservation of domain A structure.
Superimposition of three dimensional structures of domain A from *E. coli* (red), *T. thermophilus* (cyan), *S. cerevisiae* (blue), *D. melanogaster* (purple) and *H. sapiens* (orange).

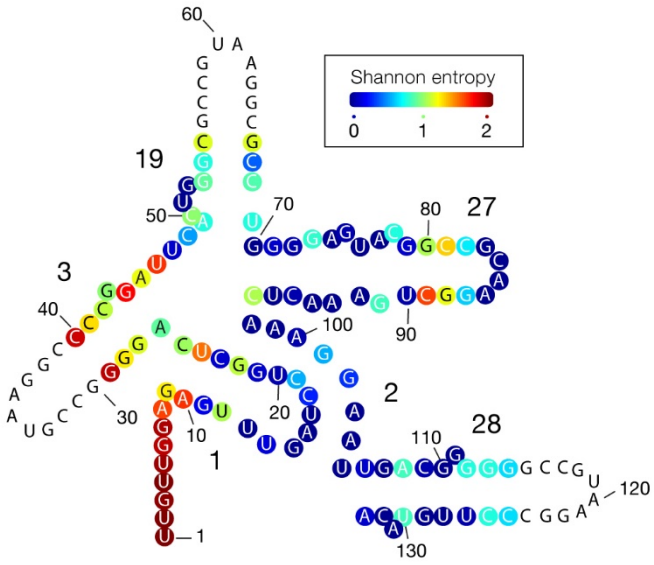

Figure 4.4 Shannon entropy data is mapped on domain A rRNA secondary structure map.
Dark blue circles indicate universally conserved nucleotides while red circles indicate variability. Nucleotide and helix numbers are shown in the figure.

and to various extents (Chen, Sim, Wurtmann, Feke, & Wolin, 2014; Felden, Florentz, Giegé, & Westhof, 1996; Fujiwara, Ito, & Nakamura, 2001; Giegé, Frugier, & Rudinger, 1998; Hirokawa, Kiel, Muto, & Selmer, 2002).

The mimicry is observed in the general L-shape, or more specifically in the amino acid acceptor stem, or in the anti-codon stem. Domain A shows some elements of tRNA mimicry (Figure 4.5). This mimicry is found in the arrangement and local conformations of helices 1, 2, and 27. Helices 1 and 2 are coaxial, and are at right angles to helix 27, giving a L-shape structure. Helix 27 of domain A is a close approximation of the anti-codon stem loop. In this region domain A is an extremely acute mimic of valine tRNA, with correct position of each base of the CAA anticodon. However, a significant difference between tRNA and domain A is seen when helix 27 is superimposed on the anticodon stem loop; helices 1 & 2

Table 4.1 Pairwise superimposition of domain A RNA.
TT is *Thermus thermophilus*. EC is *Escherchia coli*. SC is *Saccaromycies cerevisiae*. DM is *drosophila melanogaster*. HS is *Homo sapiens*.

RMSD	TT	EC	SC	DM	HS
TT					
EC	0.566				
SC	0.645	0.831			
DM	0.981	0.887	0.905		
HS	0.644	0.821	0.689	0.830	

Average RMSD = 0.827
Average RMSD = 0.778 (P atoms only)

are offset relative to the acceptor & T stems of tRNA. Ramakrishnan previously noticed a similar structural mimicry of the anticodon loop of tRNA and helix 6 of SSU rRNA (Carter et al., 2000). The 5' end of the rRNA is a rough approximation of the tRNA amino acid acceptor stem, which is formed by the 3' end of the tRNA. The relevant nucleotides of the rRNA are universally conserved (Figure 4.4) and are involved in intersubunit bridge B2c via A-minor interactions (Gao & Frank, 2006; Schuwirth et al., 2005). Where as the CCA amino acid acceptor end of the tRNA comprises a 3' OH group on its ribose sugar and can be charged by corresponding aminoacyl tRNA synthetase, domain A core rRNA contains a 5' OH group. Given the similar apparent sizes of the aboriginal structure of the LSU (Petrov et al., 2014), tRNA and domain A, this similarity may indicate the minimal size of RNA necessary to initiate complex functionality, and that larger structures were obtained only by the growth from more primitive functional structures.

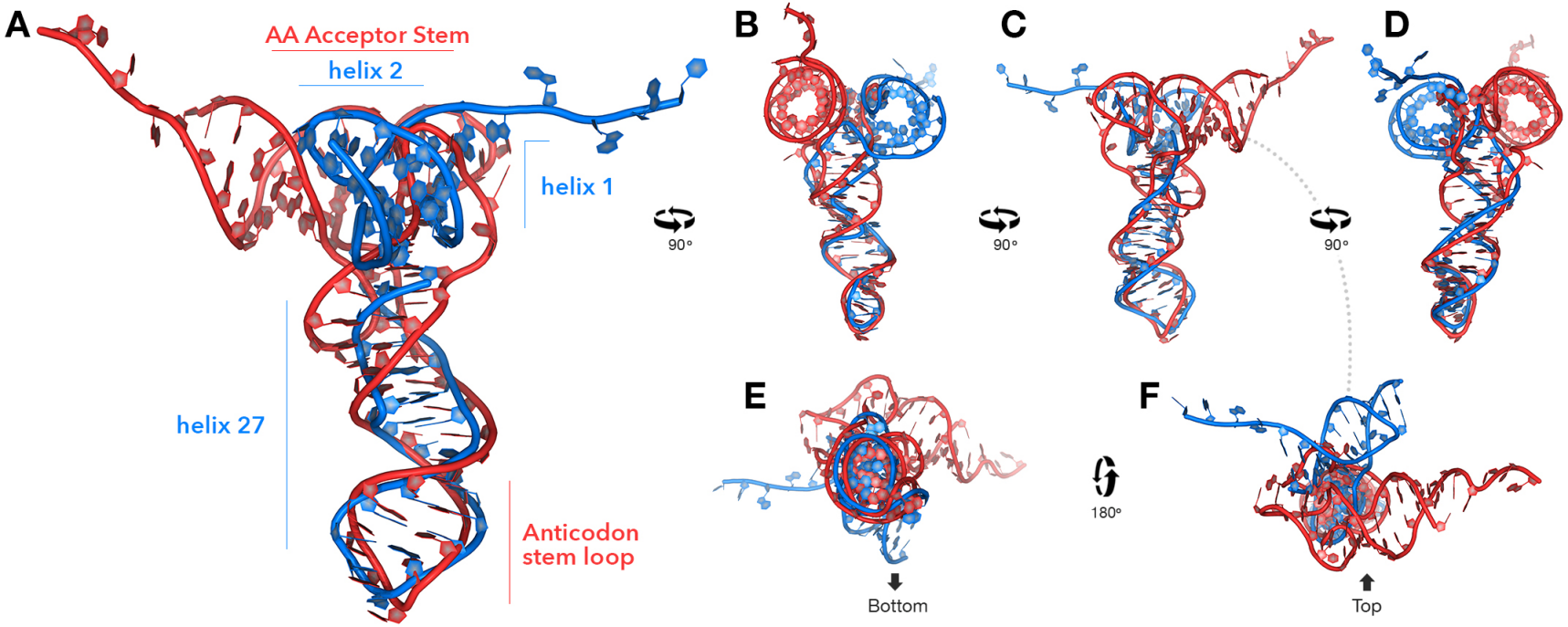

Figure 4.5 Domain A mimics tRNA.
tRNA is red and domain A is blue. (A) Superimposition of helices 1, 2, and 27 of domain A with tRNA (tRNA is from PDB ID: 2J00) (6). Both domain A and tRNA form L-shaped structures. The anticodon stem-loop superimposes on the helix 27 stem-loop. View down the CCA stem of tRNA shows that it is offset from helices 1 and 2 of domain A. (B, C, D) These views show the 90^0 rotations on y-axis. (E, F)These views show 180^0 rotations on x-axis.

4.5 Conclusion

The SSU is a central assembly of all cellular life. The architecture of the SSU has profound implications for ribosomal function and evolution. It is now possible to re-evaluate the domain structure of the SSU, which was originally established from co-variation and from low-resolution cryo-EM structures. Here, we use information from high-resolution structures for a *de novo* re-determination the domain structure of the SSU rRNA. In an analogous effort Petrov and coworkers previously re-determined the domain structure of the 23S rRNA (Petrov et al., 2013).

SSU architecture as determined from three-dimensional structures suggests peripheral rRNA domains radiate from a central core, here called domain A (Figure 4.1). The SSU therefore is dendritic in structure, in comparison to the monolithic LSU. Domain A is an autonomous core at the structural and functional center of the SSU. Domain A, which includes the central pseudoknot, is a hub that connects to the other SSU rRNA domains by helical spokes. In this model the peripheral domains of the SSU rRNA are linked to a common domain, rather than being linked directly at a common site as in the historical domain structure of the SSU rRNA.

In sum, we propose a revised architectural scheme of SSU rRNA by defining domain A, which acts as a hub to all the other domains of SSU. Moreover, domain A reveals some elements of tRNA mimicry.

CHAPTER 5

ISOLATION OF DOMAIN A

We have defined the central core of SSU RNA as a core domain of SSU, called domain A in chapter 4. In this chapter, we experimentally support our domain A model and define a structural model for isolating the domain A.

5.1 Introduction

Domain A is the core element of the SSU RNA governing the central pseudoknot and the intersubunit bridge B2c. Defined as the core domain of SSU, domain A links all the other four domains of SSU RNA functioning as an anchor. Besides the structural and architectural significance, domain A also has a functional importance in translation as discussed in earlier chapters.

To help determine if domain A meets the formal criteria of a domain, in particular the ability to autonomously fold to the native-like state, we evaluated an isolated domain A, an experimental model of domain A. We refer to this isolated domain A as "Domain A^{ISO}" (Figure 5.1).

Our selective 2'OH acylation analyzed by primer extension (SHAPE) and circular dichroism (CD) results are consistent with the autonomous folding of domain $A^{ISO.}$

We investigated the Mg^{2+}-dependence of SHAPE reactivity and CD spectra of domain A^{ISO} and several informative mutants. SHAPE and CD experiments suggest compact tertiary folding of domain A^{ISO} rRNA to a near-native state in the presence of Mg^{2+} ions. A C18A mutation in the central

pseudoknot dampens native dynamics and causes misfolding. Excision of helix 28 destabilizes of domain A^{ISO} and shifts the structure away from the native state.

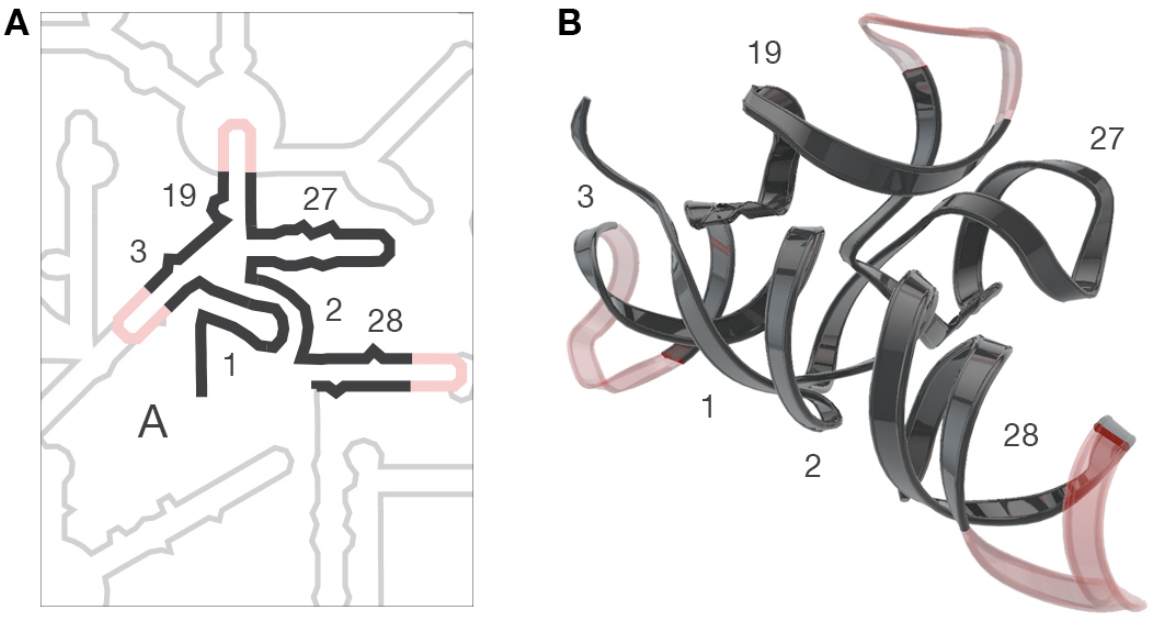

Figure 5.1 Secondary structure and three-dimensional model of domain A^{ISO}. (A) Secondary structure of domain A. Helix numbers are indicated. Domain A is black. The linkers that connect the domain A fragments to form a single RNA polymer are pink while the remainder of the 16S rRNA is grey. (B) three dimensional ribbon representation of domain A^{ISO} model colored as in panel A.

The results of experiments described here support the integrity of domain A, and our revised domain structure of the SSU rRNA. The central pseudoknot is crucial for biogenesis of the SSU, for stability of the assembled subunits, and for initiation of translation (Brink et al., 1993; Pleij et al., 1985; Poot et al., 1996; Poot, van den Worm, Pleij, & van Duin, 1998). Domain A also contains B2c, thought to be one of the oldest intersubunit bridges in the ribosome (Gao & Frank, 2006; Schuwirth et al., 2005), consistent with the hypothesis that domain A forms an ancestral core of the SSU.

5.1.1 Defining the central core by a structural model

To define the central core of SSU, we have employed domain criteria by maintaining the self-consistent molecular interactions. One goal here is to test this domain model by determining if domain A is an integrated and independent structural unit. Therefore we isolated domain A from the rest of the 16S rRNA. To form domain A^{ISO} from a single RNA polymer we linked rRNA fragments together with stem-loops (rGGCGUAAGCC) within helices 3, 19, and 28 (Figure 5.1). We characterized domain A^{ISO} and mutations and truncations of domain A^{ISO} by methods including SHAPE and CD spectroscopy.

5.2 Material and Methods

5.2.1 Chemical reagents and synthetic oligonucleotides

The chemical reagents used here are molecular biology grade or higher. DNA primers and oligonucleotides were purchased from Operon MWG. All aqueous solutions were prepared with deionized, distilled, nuclease free water (HyClone, Thermo Scientific). For the experiments with in the absence of divalent cations, nuclease free water was treated with the Chelex 100 Resin (Biorad) chelating resin and recovered with 0.2μm Ultrafree–MC–GV Centrifugal Filters (Milipore).

5.2.2 Construction of the transcription vector for domain A^{ISO} RNA

The *Thermus thermophilus* HB8 strain 16S rRNA sequence was obtained from NCBI database. The Domain A^{ISO} gene minus helix 28 was created by recursive PCR using the four oligonucleotides (5' to 3'):

Forward 1:

GGTGTGGGAATTCTAATACGACTCACTATAGGGTTGTTGGAGAGTTTGATCCTGGCT

Reverse 2:

CAGTGAATCCGGGGCCTTACGGCCCCTGAGCCAGGATCAAACTCTCCAAC

Forward 3:

GTAAGGCCCCGGATTCACTGGGCGCCGTAAGGCGCCTGGGGAGTACGGCC

Reverse 4:

CACCAAGCTTATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGC

The flanking primers were

Forward: TGAGTCGTATTAGAATTCCCACACC

Reverse: GAAACTCAAAGGAATAAGCTTGGTG.

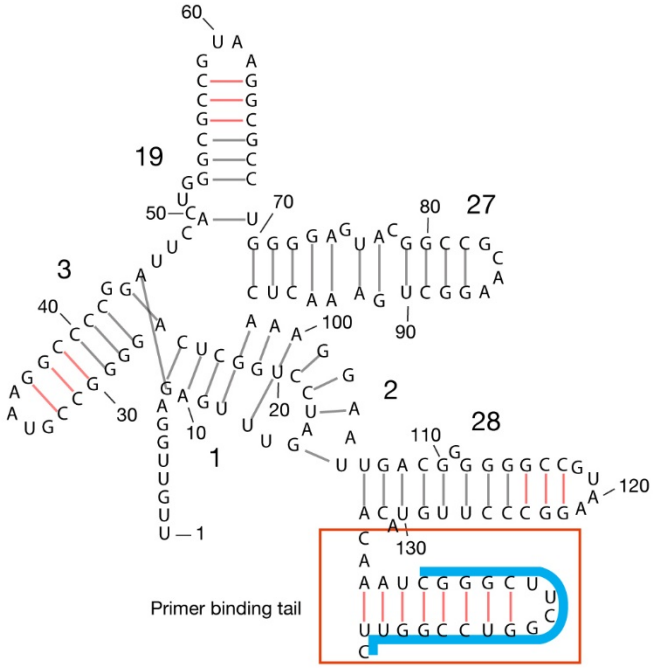

Figure 5.2 Secondary structure of domain A^{ISO} construct for SHAPE reactions. Helix and nucleotide numbers are indicated. Primer binding tail is shown in the red box and primer binding sequence is illustrated with blue line.

The domain A^{ISO} gene was cloned into the pUC19 vector using the EcoRI and HindIII restriction sites. The transformation used 5 μL of the ligation mix, which was added to 50 μL DH5α cells using the heat-shock method. Plasmids obtained by minipreps were sequenced bidirectionally by Operon MWG.

Helix 28 was added with Q5 site-directed mutagenesis (NEB) using forward AAGCTTGGCGTAATCATGG and reverse TGTACAAGGGCCTTACGG primers. The C18A mutant was also made by Q5 site-directed mutagenesis, using forward AGAGTTTGATACTGGCTCAGG and reverse CCAACAACCCTATAGTGAG primers. For SHAPE experiments, a primer binding tail was added to the 3' end by PCR using the reverse primer CACCAAGCTTGAACCGGACCGAAGCCCGATTTGTGTACAAGGGCCTTACGGC CCCCCGTCAATTCCTTTGAGTTTCAGCCTTGC (5' to 3'). The secondary structure of Domain A^{ISO} with the SHAPE tail is shown in the Figure 5.2.

5.2.3 *In vitro* transcription and purification of domain A^{ISO} RNA

The pUC19 plasmid containing the domain A^{ISO} gene was digested with HindIII-HF (NEB) for 2 hours at 37 °C as otherwise described by manufacturer. The reaction mixture was incubated at 80 °C for 20 minutes to deactivate the enzyme. The reaction was purified with SmartSpin nucleic acid & purification columns (Denville Scientific Inc.) using DNA Clean & Concentrator Kit buffers (Zymo Research Corp.) Digested plasmid (400-1000 ng) was used as a template for T7 RNA polymerase (NEB) transcription. Run-off transcription reaction was prepared according to manufacturer's description (NEB T7 High Yield RNA Synthesis Kit). The reaction mixture was incubated for 16 hours at 37 °C. After incubation, 1 μL Turbo DNAse (Ambion) was added to the reaction mixture,

which was then incubated for 30 minutes at 37 °C. The RNA was purified by ammonium acetate precipitation. Ultimately, 40 μL nuclease-free H_2O was added to the dried pellet and the OD was measured with a Nanodrop (Thermo Scientific). The RNA was further purified by G25 size exclusion chromatography (illustra™NAP™-10, GE Healthcare).

5.2.4 Selective 2'OH acylation analyzed by primer extension reactions and dideoxy sequencing

5.2.4.1 Dideoxy sequencing reactions

Dideoxy sequencing reactions were carried out by heating a 20 μL solution of 50 ng/μL domain A^{ISO} rRNA mixed with 10 μL (0.8 μM) 5' 6-FAM labeled GAACCGGACCGAAGCCCG primer (Operon MWG). To anneal the primer to the RNA, the reaction was heated to 85 °C and slowly cooled to 30 °C at a rate of 1.5 °C per minute. For domain A^{ISO} rRNA lacking helix 28, 5' 6-FAM labeled TATTCCTTTGAGTTTCAGCC primer was used. After primer annealing, 20 μL mixture of SuperScript® III Reverse Transcriptase (Invitrogen) reaction mixture prepared and added to domain A^{ISO} rRNA and primer mixture to give the final concentrations of 1X RT buffer, 2 mM DTT, 0.625 mM dNTPs, and 2.5mM ddNTPs (TriLink BioTechnologies). The reverse transcription reaction was carried out by incubating 50 μL reaction mixture at 55 °C for 2 hours and quenched for 15 min by heating to 70 °C.

5.2.4.2 Selective 2'OH acylation anayzed by primer extension reactions

For the SHAPE reactions, a 70 μL solution of 150 ng/μL domain A^{ISO} rRNA was incubated for 4 min at 85 °C in the presence of 5 μM 1,2-

diaminocyclohexanetetraacetic acid (DCTA) (Sigma) chelating agent and allowed to cool for 10 min at room temperature. This procedure depletes divalent cations from the RNA. Divalent-free RNA was divided to two 32 μL samples and 4 μL of 10X folding buffer was added (500mM HEPES pH 8.0, 2.5 M NaCl) for RNA folding with sodium. Four μL 10X folding buffer was added (500mM HEPES pH 8.0, 2.5 M NaCl, 20mM $MgCl_2$) for RNA folding with sodium and magnesium. The sample was folded by incubated at 20 min at 37 °C and was divided two 18 μL solutions. One of the solutions was added to 2 μL of 800 mM benzoyl cyanide in anhydrous DMSO. The other solution was added 2 μL of pure DMSO for a negative background control. The reaction mixture was incubated 2 min at room temperature. The modified RNA was purified using Zymo RNA Clean and Concentrator Kit and eluted in 25 μL modified TE buffer (10mM Tris, 0.1 mM EDTA). Primer annealing and extension reactions were as described above.

5.2.4.3 Capilarry electrophoresis

For the capillary electrophoresis, 1.5 μL of reverse transcription reaction mixture was mixed with 0.5 μL ROX-labeled DNA sizing ladder and 9 μL of HiDi Formamide (Applied Biosystems) in a 96-well plate. To denature the cDNA, the plate was incubated for 5 min at 95 °C. The mixture was resolved on a 3130 Genetic Analyzer (Applied Biosystems). Capillary electrophoresis data were processed using in-house MatLab scripts as described (Athavale, Gossett, et al., 2012). First, data were aligned via standard peaks and the baseline was subtracted. Sequencing peaks were matched with SHAPE data peaks. The traces were integrated and processed with a signal decay correction, and were scaled and normalized.

5.2.5 Circular dichroism spectroscopy

A solution of 25 ng/mL RNA, 5mM sodium cacodylate, pH 6.8 was titrated with either a EDTA or Mg^{2+}. The RNA was titrated first with the chelator, followed by back-titration with Mg^{2+}, taking CD scans on a Jasco J-810 spectropolarimeter after each addition. Four CD spectra averaged, from 350 to 220 nm with an integration time of 4 seconds, bandwidth of 4 nm, a scan speed of 50 nm/min. The temperature was kept at 20 °C.

5.2.6 3D modelling of domain A^{ISO}

The domain A three-dimensional structure was modeled from assembled ribosome structure of Ramakrishnan (PDB ID:2J00) (Selmer et al., 2006). Nucleotides 1-29, 554-569, 881-929, and 1388-1396 were extracted from the crystal structure and capped by a stem-loop containing the three base pairs and a tetra loop with a sequence GCCGUAAGGC. The 3D coordinates of the stem loop were extracted from Hsiao (Hsiao et al., 2013). The stem loops were positioned as extensions of the Domain A helices and connected to it by adding the 3'-P bonds. The stem loops along with their two adjacent base pairs from the Domain A were subjected by the partial energy minimization, while the rest of the structure was held fixed.

5.2.6.1 Energy minimization

Partial minimization of the re-ligated rRNAs was performed with Sybyl-X 1.2 software (Tripos International, St. Louis, MO, USA) with the AMBER FF99 force field using an implicit solvent model with the distance dependent dielectric function D(r) = 20r. The non-bonded cut-off distance was set to 12 Å. Each system was

minimized by 1000 steps of steepest decent followed by 5000 steps of conjugate gradient minimization.

5.2.7 Data mapping

SHAPE data are normalized and mapped on in-house RiboVision server using the custom data function (Bernier et al., 2014).

5.2.8 Figures and images

Figures of three-dimensional structures are prepared with PyMol or Maxon Cinema 4D with the ePMV plugin (Johnson, Autin, Goodsell, Sanner, & Olson, 2011). Secondary structures are obtained from in-house RiboVision server (Bernier et al., 2014). Labels are added in Adobe Illustrator or Adobe Photoshop.

5.3 Results and Discussion

5.3.1 *In vitro* folding of domain A

We have asked if domain A is an interdependent and integrated structural unit, satisfying the criteria for a domain, by assaying the SHAPE reactivity and CD spectra of domain A^{ISO} in the absence and presence of Mg^{2+} ions. We test the domain A model by predicting the effects of mutations on domain A^{ISO}. We compare the SHAPE reactivity of domain A^{ISO} with that of the same rRNA elements within the intact SSU, previously published by Weeks and coworkers (McGinnis & Weeks, 2014). Three-dimensional and secondary structures can be probed with SHAPE because paired nucleotides in double-stranded regions are less reactive to the SHAPE reagent than unpaired nucleotides in loops, bulges and single strands (Wilkinson, Merino, & Weeks, 2006).

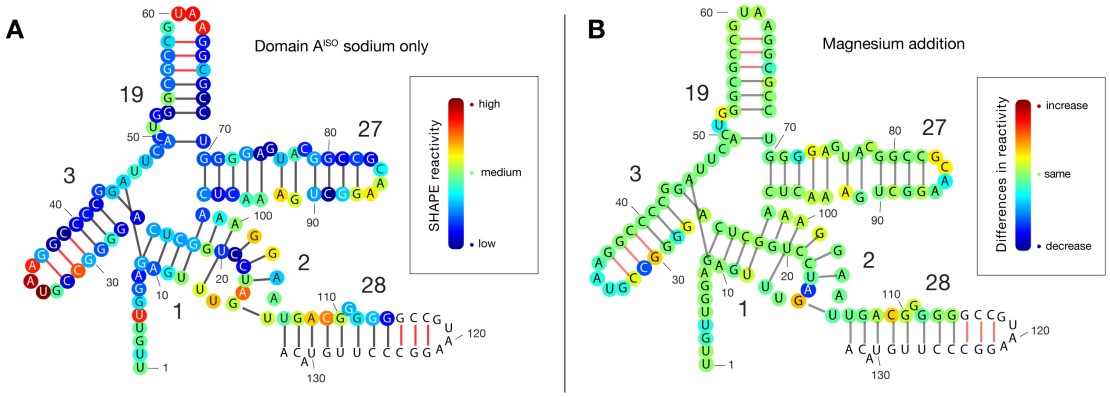

Figure 5.3 SHAPE reactivity of domain A^{ISO} mapped onto the proposed secondary structure in the presence of sodium and magnesium cations.
Base pairs predicted from the secondary structure of the intact SSU are indicated by black lines. Presumed base pairs in the linkers are indicated by red lines. Helix and nucleotide numbers are indicated. (A) Reactivity of domain A^{ISO} in the presence of Na^+ only (250 mM). The red circles indicate high reactivity while the blue circles indicate low reactivity. The color scale is shown in the outbox. (B) Difference in SHAPE reactivity upon addition of Mg^{2+} (2 mM). Red indicates a increase in reactivity, while blue indicates a decrease. Green indicates no change. The coloring scheme is shown in the outbox. Data were not accessible for the uncolored nucleotides. The primer binding tail is omitted for clarity. The full sequence of the construct is shown in

Nucleotides involved in tertiary and Mg^{2+} interactions change reactivity upon the addition of Mg^{2+} (Athavale, Gossett, et al., 2012; Athavale, Petrov, et al., 2012; Hsiao et al., 2013; Mortimer & Weeks, 2007).

The SHAPE data suggest that in the presence of Na^+ alone, domain A^{ISO} folds to form helices 1, 2, 3, 19, 27 and 28 (Figure 5.3). For helices 1, 3 and 19, the duplex regions are unreactive and the loop regions are reactive. High reactivity of C31 suggests a defect near the loop of helix 31. Helix 27 shows the same anomalous pattern of reactivity in domain A^{ISO} as in the intact SSU (Figure 5.4). Helices 2 and 28 are anomalously reactive in domain A^{ISO}, consistent with their anomalous reactivity in the intact SSU (Deigan et al., 2009; McGinnis & Weeks,

2014). Nucleotides involved in base triples in the intact SSU (nucleotides G9, U20, and G22) show suppressed reactivity in domain A^{ISO}. The 5' terminus of domain A^{ISO} (which is also the terminus of the 16S rRNA) shows elevated SHAPE reactivity as expected of unstructured RNA. Similarly, the single-stranded nucleotides between stems 3 and 19 (A45, U46, U47) have higher reactivity than the flanking stems.

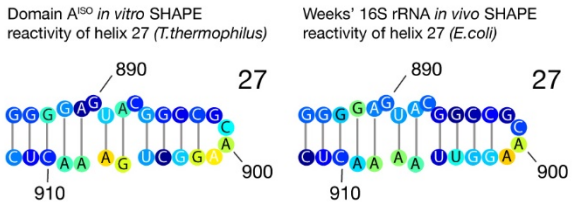

Figure 5.4 Comparison of SHAPE reactivity between domain A^{ISO} helix 27 and Weeks' intact 16S rRNA helix 27.
Domain A^{ISO} helix 27 *in vitro* SHAPE reactivity from *T.thermophilus* is consistent with Weeks' *in vivo* 16S rRNA from *E.coli*. Nucleotide numbers are following *E.coli* numbering scheme. Base pairs are indicated with black line.

Mg^{2+} ions appear to stabilize domain A^{ISO} and facilitate full folding to the native state. Monovalent cations generally allow RNAs to form secondary structures and a subset of tertiary interactions. Divalent cations are required for complete folding to the native state (Bowman, Lenz, Hud, & Williams, 2012; Woodson, 2010). Here we used CD spectroscopy to characterize the effects of divalent cations on the structure of domain A^{ISO} (Figure 5.5). The addition of Mg^{2+} to the Na^{+} form of domain A^{ISO} increases the intensity of the diagnostic CD band at 265 nm. The intensity increases over the range of [Mg^{2+}] from 0 to 700 μM after which it plateaus. These results show that the Mg^{2+} effects on domain A^{ISO} are similar to those of well-characterized RNAs such as tRNA (Römer & Hach, 1975)

and P4-P6 of the tetrahymena group I ribozyme (Frederiksen, Li, Das, Herschlag, & Piccirilli, 2012).

The CD results are consistent with SHAPE reactivity. Mg^{2+} has subtle but widely distributed effects on the SHAPE reactivity of domain A^{ISO}. Mg^{2+} is expected to influence the SHAPE reactivity of nucleotides that directly contact

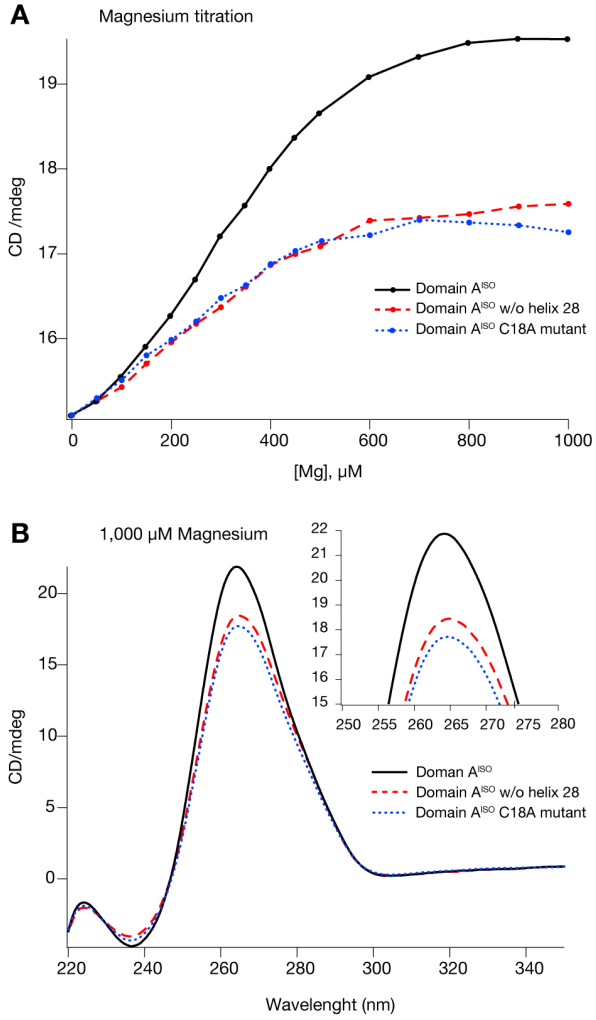

Figure 5.5 Circular dichroism spectroscopy of domain A^{ISO}.
(A) Mg^{2+} titration of domain A^{ISO} rRNA (solid black), the C18A mutant of domain A^{ISO} (dashed blue), and domain A^{ISO} rRNA with helix 28 excised (dotted red). Mg^{2+} concentration is plotted versus the intensity of the diagnostic CD peak (265 nm). (B) CD spectra of the same series of RNAs in the presence of 1.0 mM Mg^{2+}. The outbox shows a close-up of the 265 nm peak. Initial rRNA samples were depleted in Mg^{2+} ions.

Mg^{2+} or are involved in Mg^{2+}-dependent tertiary interactions. This pattern of Mg^{2+}-dependent SHAPE reactivity has previously been observed for tRNA, RNase P, the P4-P6 domain of the *Tetrahymena* Group I intron and Domain III of the 23S rRNA (Athavale, Gossett, et al., 2012; Athavale, Petrov, et al., 2012; Hsiao et al., 2013; Merino, Wilkinson, Coughlan, & Weeks, 2005; Mortimer & Weeks, 2008). Nucleotides in domain A^{ISO} overall show slight decreases in SHAPE reactivity while some loop regions and bulges show increases (Figure 5.3). Reactivity of nucleotides A16 and C31 drop upon addition of Mg^{2+} suggesting that correct folding of Helix 3 requires Mg^{2+}. Based on the intact SSU, A16 is expected to interact directly with a Mg^{2+} ion in the native structure (Selmer et al., 2006). Indeed, A16 shows the greatest change in SHAPE reactivity of any site in domain A^{ISO} upon addition of Mg^{2+}.

5.3.2 Helix 28 is an integral component of domain A

The central pseudoknot is formed by helices 1 and 2 (Petrov et al., 2014) (Figure 5.1). We anticipate that the structure and stability of the central pseudoknot, and of domain A^{ISO} should be dependent on helix 28, because it forms a continuous stack with helix 2 in the intact SSU (Chapter 4, Figure 4.2). If our model of domain A^{ISO} is correct, helix 28 contributes globally to the stability of domain A^{ISO}. Therefore, we have determined the effect of excision of helix 28 from domain A^{ISO}.

Upon excision of helix 28, changes in SHAPE reactivity are distributed throughout domain A^{ISO} consistent with global changes in structure (Figure 5.6). Reactivity increases near the 5' terminus. Within helix 1, increases in SHAPE

reactivity suggest disruption of base pairs G9-C25, A10-U24, G11-C23, U12-G22, and U13-U20 (Figure 5.6).

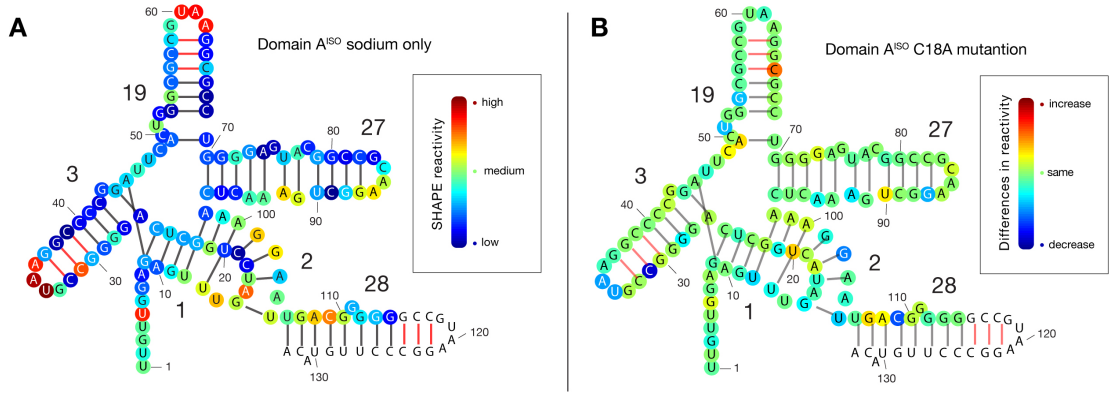

Figure 5.6 SHAPE reactivity of the intact domain A^{ISO} and helix 28 excised mutant. Base pairs predicted from the secondary structure of the intact SSU are indicated by black lines. Presumed base pairs in the linkers are indicated by red lines. Helix and nucleotide numbers are indicated. (A) Reactivity of domain A^{ISO} in the presence of Na^+ only (250 mM). The red circles indicate high reactivity while the blue circles indicate low reactivity. The color scale is shown in the outbox. (B) Difference in SHAPE reactivity upon excision of helix 28. Red indicates a increase in reactivity, while blue indicates a decrease. Green indicates no change. The coloring scheme is shown in the outbox. Data for truncated domain A^{ISO} are acquired in the presence of both Na^+ and Mg^{2+}. Data were not accessible for the uncolored nucleotides. The primer binding tail is omitted for clarity.

Furthermore, it appears that base pairing is precluded between U14 and A16 in both the intact SSU (Petrov et al., 2014; Selmer et al., 2006) and in domain A^{ISO}. These nucleotides are in a loop region in the native structure, and show a higher SHAPE reactivity than other sites in the central pseudoknot (Figure 5.6A). However, when helix 28 is omitted from domain A^{ISO}, U14 and A16 decrease in reactivity (Figure 5.6), suggesting formation of non-native pairing interactions.

Support for the significance of helix 28 in domain A^{ISO} structural integrity is provided by CD spectroscopy. The effect of Mg^{2+} on the CD spectrum of domain A^{ISO} is diminished by excision of helix 28. Figure 5.5A demonstrates that changes

in CD spectra after addition of Mg^{2+} are lessened by approximately 50% for domain A^{ISO} lacking helix 28 compared to intact domain A^{ISO}. The diagnostic 265 nm peak does not reach full intensity in the absence of helix 28. Combined SHAPE and CD data suggest that formation of the native folded state of domain A^{ISO} is dependent of helix 28, supporting the revised SSU domain model.

5.3.3 Disruptive mutations in the central pseudoknot prevent domain A^{ISO} folding to the native state

Dahlberg suggested that helix 27 is dynamic (Lodmell & Dahlberg, 1997). This model is experimentally supported by cryo-EM of Frank and coworkers (Gabashvili et al., 1999). The unusually high SHAPE reactivity throughout helix 27, in both domain A^{ISO} (here) and in the intact SSU (Weeks (McGinnis & Weeks,

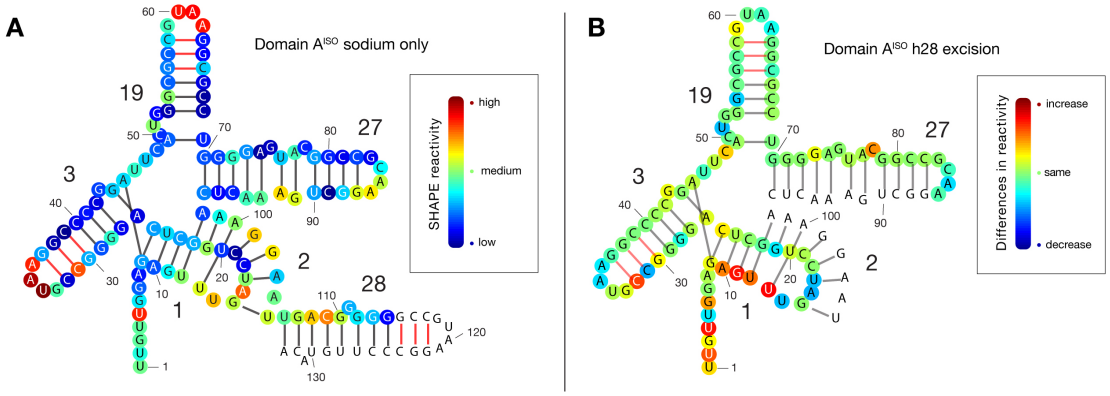

Figure 5.7 SHAPE reactivity of the intact domain A^{ISO} and C18A mutant. Base pairs predicted from the secondary structure of the intact SSU are indicated by black lines. Presumed base pairs in the linkers are indicated by red lines. Helix and nucleotide numbers are indicated. (A) Reactivity of domain A^{ISO} in the presence of Na^+ only (250 mM). The red circles indicate high reactivity while the blue circles indicate low reactivity. The color scale is shown in the outbox. (B) Difference in SHAPE reactivity upon mutation of C18 to A18. Red indicates a increase in reactivity, while blue indicates a decrease. Green indicates no change. The coloring scheme is shown in the outbox. Data for mutant domain A^{ISO} are acquired in the presence of both Na^+ and Mg^{2+}. Data were not accessible for the uncolored nucleotides. The primer binding tail is omitted for clarity.

2014)) is consistent with the Dahlberg model. Pleij, Brink and coworkers demonstrated that a C18A mutation within the central pseudoknot has significant effects *in vivo*, inhibiting translation by affecting subunit assembly (Brink et al., 1993; Poot et al., 1998). This mutation is expected to disrupt the C18-G102 base pair.

To test the Brink and Dahlberg hypotheses and our model of domain A^{ISO}, we constructed domain A^{ISO} containing the C18A mutation. We observe that domain A^{ISO} is significantly affected by this mutation (Figure 5.7). The C18A mutation lowers the general SHAPE reactivity of the domain A^{ISO}, extending even to the 5' single stranded region (Figure 5.7). Thus the mutation appears to dampen mobility. More specifically, the C18A mutation intensifies the reactivity of C18 and reduces the reactivity of G102. Similarly, the C18A mutant affects the CD spectra of domain A^{ISO}. The mutation modulates the effect of Mg^{2+} on the intensity of the 265 nm band (Figure 5.5). These results suggest that this mutation, like the excision of helix 28, prevents native folding of domain A^{ISO}. In sum, the data appear to support the models of Dahlberg and Brink, and our domain model, incorporating domain A. The C18A mutation and excision of helix 28 give similar effects in the Mg^{2+}-dependence of the CD spectra, which are indications of misfolding of domain A^{ISO}.

5.4 Conclusion

We have constructed an experimental system for Domain A and our data suggests that domain A meets the formal criteria of a domain, in particular the ability to autonomously fold to the native-like state. We investigated the Mg^{2+}-dependence of SHAPE reactivity and CD spectra of domain A^{ISO} and several

informative mutants. SHAPE and CD experiments support compact tertiary folding of domain A^{ISO} rRNA to a near-native state in the presence of Mg^{2+} ions. Mutation experiments provide valuable information for the central pseudoknot. A C18A mutation in central pseudoknot disrupts native dynamics and causes misfolding. Exclusion of helix 28 destabilizes domain A^{ISO} and disrupts the structure of the native state. The results of experiments described here support the integrity of domain A, and our revised domain structure of the SSU rRNA. The central pseudoknot plays a significant role in biogenesis of the SSU, stabilizing the assembled subunits, and initiating the translation (Brink et al., 1993; Pleij et al., 1985; Poot et al., 1996; Poot et al., 1998). These results are consistent with the hypothesis that domain A forms the core of the SSU and has the ability to fold independently into near native state.

CHAPTER 6

INTERACTIONS OF DOMAIN A WITH RIBOSOMAL PROTEINS

The binding of ribosomal proteins to rRNA is a hierarchal process. Nomura suggested an assembly map for 30S small subunit protein binding hierarchy (Mizushima & Nomura, 1970). In chapter 1, we have briefly mentioned the small subunit proteins and their binding locations as well as their binding hierarchy. In this chapter, we will focus on the small subunit proteins, which bind to domain A^{ISO} RNA. Our goal is to investigate the binding of ribosomal proteins S5 and S12 to domain A^{ISO} RNA in that these proteins interact with the 16S rRNA in native structure.

6.1 Introduction

Domain A forms the central core of the SSU rRNA. In the native state, ribosomal proteins S5 and S12 interact with the domain A region of the SSU RNA. A recent 30S assembly map of Williamson and coworkers suggests that S5 and S12 are not primary binding proteins (Mulder et al., 2010). In this connection, their specific binding to the 16S rRNA would require primary and secondary proteins to bind first. Here we tested the binding of S5 and S12 to domain A^{ISO} RNA *in vivo* and *in vitro*.

6.1.1 Tertiary binding proteins

Ribosomal proteins S5 and S12 are tertiary binding proteins. S5 requires prior binding of S4, S8, and S20 as primary binding proteins and S16 as a secondary binding protein. S12 requires S4, S20, S17 and S8 as primary binding

proteins and S16 and S5 as secondary and tertiary binding proteins. Figure 1.5 of chapter 1 illustrates the binding hierarchy.

6.2 Materials and Methods

6.2.1 Yeast-three hybrid assay

6.2.1.1 Preparation of the clonning vectors

6.2.1.1.1 Cloning vector for domain A^{ISO}

The domain A^{ISO} gene was constructed with recursive PCR under standard cycling conditions by using primers (5' to 3'):

Forward 1:

"GGTGTGGGAATTCTAATACGACTCACTATAGGGTTGTTGGAGAGTTTGATCCTGGCT"

Reverse 2:

"CAGTGAATCCGGGGCCTTACGGCCCCTGAGCCAGGATCAAACTCTCCAAC"

Forward 3:

"GTAAGGCCCCGGATTCACTGGGCGCCGTAAGGCGCCTGGGGAGTACGGCC"

Reverse 4:

"CACCAAGCTTATTCCTTTGAGTTTCAGCCTTGCGGCCGTACTCCCCAGGC"

Flanking primer forward: "TGAGTCGTATTAGAATTCCCACACC"

Flanking primer reverse: "GAAACTCAAAGGAATAAGCTTGGTG"

The PCR product was cloned to the pIIIA/MS2-1 vector from SphI restriction site. The plasmid containing domain A was confirmed by bidirectional sequencing by

MGW Operon. Helix 28 of domain A was added by NEB Q5 site-directed mutagenesis as described by the manufacturer (NEB) by using the primers:
"AAGGCCCTTGTACAGCATGCAAGCTGCCCGGG"
"ACGGCCCCCCGTCAATTCCTTTGAGTTTCAGCCTTGCGG".

6.2.1.1.2 Cloning vector for S5 and S12

The gene sequences of *T. thermophilus* HB8 ribosomal proteins S5 and S12 are obtained from the NCBI database.
Ribosomal proteins S5 and S12 genes are amplified from *T. thermophilus* HB8 genomic DNA by using primers (5'to3'):
S5 forward: "GGTGGAATTCCCGGAGACCGACTTTGAAGA"
S5 reverse; "CCCACTCGAGACCTTGAGCCTGGGCATG"
S12 forward: "GTGGCCCGGGGTGGTGGCACTGCCGACG"
S12 reverse: CCCAGAATTCCTTCTTGGCCGCGGTCTTG
The PCR products are cloned into the pACT2 vector from EcoRI and XhoI sites for S5 and EcoRI and XmaI sites for S12.

6.2.1.2 Yeast transformations

A freshly streaked yeast strain YBZ-1 was inoculated in 2 mL autoclaved YPAD media (yeast extract (10g/L), peptone (20g/L), glucose (20g/L), adenine hemisulfate (400mg/L)) and incubated 16 hours at 300 rpm at 30 °C.

The overnight culture was pelleted at 13000 RPM on a tabletop centrifuge for 30 sec. The transformation mixture was added to the pellet in the order:
PEG 3500 50% v/v (240 μL), LiAc 1M (36 μL), boiled salmon sperm DNA (50 μL), plasmid DNA 800ng + co-activator plasmid (800ng) + dH_2O (34 μL).

The transformation mix was vortexed and incubated in a water bath at 42 °C for 1 hour. After incubation, the transformation mix was centrifuged 30 sec at top speed and the supernatant was removed by pipetting. The pellet was resuspended with 1 mL sterile water by pipetting up and down. After resuspension, 200 μL of the transformation mix was plated on CM-AL plates (adenine and leucine deficent complete media / CM: YNB (1.5 g/L), ammonium sulfate anhydrous (5 g/L), dextrose (20 g/L), dropout powder (1.3 g/L) (L-arginine (HCl) (40 μg/mL), L-alinine (20 μg/mL), L-aspartic acid (100 μg/mL), L-asparagine (40 μg/mL), L-cystene (40 μg/mL), L-glutamic acid (100 μg/mL), L-glutamine (40 μg/mL), L-glycine (40 μg/mL), L-lysine (monoHCl) (30 μg/mL), L-methionine (20 μg/mL), L-phenylalanine (50 μg/mL), L-serine (375 μg/mL), L-proline (40 μg/mL), L-threonine (200 μg/mL), L-tyrosine (30 μg/mL), L-valine (150 μg/mL) L-methionine (100 μg/mL)), histidine (20 μg/mL), tryptophan (40 μg/mL), uracil (20 μg/mL) and incubated at 30 °C for 3-4 days. Picked single colonies were confirmed to contain the double transformation by colony PCR.

6.2.1.3 Absorbance measurements

Confirmed colonies were grown in CM-AL media overnight and diluted with CM-ALH (same as CM-AL with additional histidine deficiency) and 20 μL diluted cultures were added to a 96-well plate. Increasing concentrations of 3AT(3-amino triazole) were added to the media as an inhibitor of histidine production pathway (0.2 μM, 0.4 μM, 0.6 μM, and 0.8 μM). The OD was read with the BioTek Synergy H4 hybrid reader in 24 hour timeframes for a total of 72 hours.

6.2.1.4 Overview of yeast-three hybrid assay

The yeast-three hybrid system is a powerful tool to monitor protein - RNA interactions *in vivo*. Association of the target protein and the target RNA activates the histidine reporter gene that enables the histidine production pathway in yeast. Observation of the cell growth in absence of histidine in the media reveals the binding between the target RNA and the target protein. A competitive inhibitor of the histidine production pathway assists the quantitative

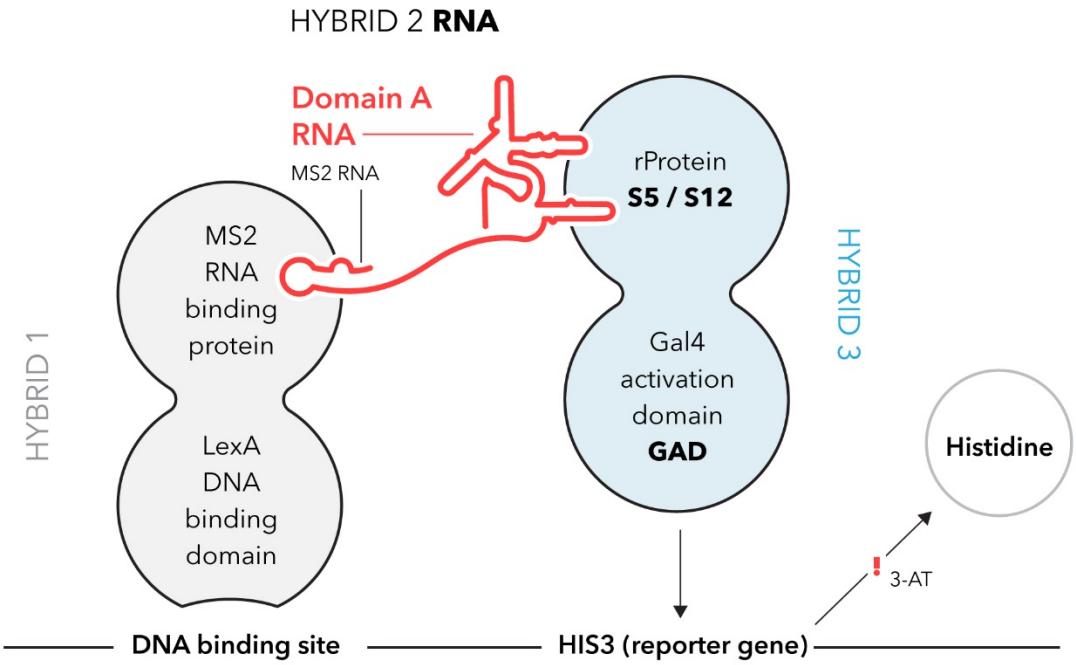

Figure 6.1 Schematic representation of Y3H system for determining RNA-protein interactions.
The LexA operator in *Saccharomyces cerevisiae* strain YBZ-1 controls the HIS3 reporter gene. Hybrid 1 is composed of LexA/MS2 coat protein fusion which binds to the LexA operator. The MS2 coat protein interacts strongly to the MS2 sequence of the hybrid RNA, which is designed to carry the MS2 RNA and RNA sequence of interest, e.g., domain A^{ISO} RNA. In Hybrid 3, the yeast GAL4 transcriptional activation domain (GAD) is attached to the protein of interest (e.g., rProteins S5 or S12). The HIS3 reporter gene is expressed as a result of *in vivo* binding of RNA of interest and protein of interest which completes hybrid 2. The affinity of RNA-protein binding is evaluated by resistance to a competitive inhibitor of the HIS3 product which is 3-Amino-1,2,4-triazole (3-AT), imidazoleglycerol-phosphate dehydratase.

analysis of the binding affinity. A schematic demonstration of the yeast-three hybrid assay is illustrated in Figure 6.1.

6.2.2 Electrophoretic mobility shift assay

Domain A^{ISO} RNA was added the binding buffer (25 mM HEPES pH 8.0, 5% glycerol, 10 mM $MgCl_2$, 20 mM KCl, and 0.1 mM EDTA) and heated to 80 °C for 4 min and cooled on the bench at room temperature for 10 min. Ribosomal protein S5 or S12 fusion was added to the mixture and the total reaction was 10 μL. The reaction mixture was incubated at room temperature for 45 min and the loading buffer was added in final concentration of 1X. The 5X loading buffer is composed of 50 mM Tris-HCl (pH 6.8), 0.01 % (w/v) bromophenol blue, 25 % (v/v) glycerol. Six percent native polyacrylamide gel was prepared with 2 mL 40% acrylamide (acrylamide:bis-acrylamide, 29:1) (Fisher Bioreagents), 2.5 mL 4X Tris-Glycine buffer (100mM Tris-HCl (pH 8.5) and 800 mM Glycine), 5.4 mL ddH_2O (Hyclone, Thermo Scientific), 100 μL 10% APS (ammonium persulfate), and 8 μL TEMED (N,N,N,N- tetramethyl-ethylenediamine) (National diagnostics). The running buffer is composed of 25 mM Tris and 200 mM Glycine (pH 8.5). The gel was run for 50 min at 120 V at room temperature.

The gel then was stained with SYBR Green I Nucleic Acid Gel Stain as described by the manufacturer (Lonza) for 20 min. After first staining, gel then was incubated in 7.5 % acetic acid for 20 min. After acetic acid incubation, the gel then was stained with SYBRO Red Protein Stain (Lonza) for 45 min. In each reaction 880 ng domain A^{ISO} RNA was used as a fixed amount while the protein S5 or S12 fusion amount varied from a 1:1 ratio to a 1:16 ratio. The stoichiometric

ratio was calculated as 880 ng RNA to 1200 ng of the S5 fusion protein or 1146 ng of the S12 fusion protein for a 1:1 ratio.

6.2.3 *In vitro* folding studies

A six percent native polyacrylamide gel was prepared with 2 mL 40% acrylamide (acrylamide:bis-acrylamide, 29:1) (Fisher Bioreagents), 2.5 mL 4X Tris-Glycine buffer (100mM Tris-HCl (pH 8.5) and 800 mM Glycine), 5.4 mL ddH_2O (Hyclone, Thermo Scientific), 100 μL 10% APS, and 8 μL TEMED (National diagnostics). The native PAGE solution mixture was poured in 1mm glass. Domain A^{ISO} RNA folding was performed by incubating the RNA at 90 °C for 3 min and cooling to room temperature for 10 min. The 2X native PAGE dye (40mM Tris-HCl pH 6.8, 0.02% (w/v) bromophenol blue, 50% (v/v) glycerol) was added to the RNA sample and loaded into wells and run 40 min at 155 V at room temperature. The gel then was stained with SYBR Green I Nucleic Acid Gel Stain as described by the manufacturer (Lonza). In each reaction, 880 ng domain A^{ISO} RNA was used. For P4-P6 RNA, 912 ng RNA was used. For the duplex 12 mer, 330 ng RNA was used.

The duplex RNA sequence for each strand is listed below from 5' to 3': "GGUGAGGCGGUG" and the complementary strand "CACCGCCUCACC" (Operon).

The gel image was taken with a Typhoon Imager (acquisition mode: fluorescence, filters 520BP40 and 610BP30, laser 488 nm, sensitivity normal, PMT 400V, pixel size 100 microns).

6.2.4 Preparation of the cloning vectors for purification of S5 and S12

6.2.4.1 Maltose binding protein fusion

Ribosomal proteins S5 and S12 genes were amplified from *T. thermophilus* HB8 genomic DNA. For ribosomal protein S5, primers (5'to3'):

Forward: "CCGGAGACCGACTTTGAAGA"

Reverse: "ACCTTGAGCCTGGGCATG" were used for the first amplification with standard cycling conditions and re-amplified with the reverse stop codon-EcoRI primer: "CCCAGAATTCTTAACCTTGAGCCTGGGCATG" to add the stop codon and EcoRI restriction site into the S5 gene.

For the ribosomal protein S12, primers:

Forward: "GTGGTGGCACTGCCGACG"

Reverse: "CTTCTTGGCCGCGGTCTTG" were used for the first amplification and re-amplified with the reverse stop codon-EcoRI primer: "CCCAGAATTCCTACTTCTTGGCCGCGGTCTTG" to add the stop codon and EcoRI restriction site into the S12 gene. The PCR products were cloned into pMALc5x vector from XmnI and EcoRI sites as described by NEB (NEB pMAL protein prufication kit).

The cloned plasmid was transferred to BL21 *E.coli* competent cells with heat-shock transformation method.

6.2.4.2 Intein mediated protein purification

Ribosomal proteins S5 and S12 genes were amplified from *T. thermophilus* HB8 genomic DNA or a pMALc5x cloned plasmid. For the ribosomal protein S5, primers (5'to3'):

Forward: "GGTGCATATGCCGGAGACCGACTTTGAAG"

Reverse: "CCCAGCTCGAAGAGCTTAACCTTGAGCCTGGGCATGG" were used for amplification of the gene. The PCR product for ribosomal protein S5 was digested and cloned into pTXB1 vector from NdeI and SapI restriction sites as described by NEB (IMPACT Protein Purification Systems, NEB).

For the ribosomal protein S12, primers (5' to 3'):

Forward: "GGTGCATATGGTGGCACTGCCGACGA"

Reverse: "CCAACTAGTGCATCTCCCGTGATGCATTGAGCCTGGGCATGGGC" were used for the amplification. The PCR product for ribosomal protein S12 was digested and cloned to pTXB1 vector from NdeI and SpeI restriction sites for 5' and 3' ends respectively. To increase the in-column cleavage efficiency during affinity purification of ribosomal protein S12, the GGT codon, which codes for glycine was omitted and glutamate became the C-terminal residue (IMPACT Protein Purification Systems, NEB).

The cloned plasmid was transferred to BL21 *E.coli* competent cells with the heat-shock transformation method

6.2.5 Purification of S5 and S12

6.2.5.1 Amylose affinity column

BL21 cells containing pMALc5x-S5 or pMALc5X-S12 were grown in ZYM-505 media for 16 hours at 37 °C ((Studier, 2005)). Starter culture was inoculated in ZYM-5052 auto induction media and was grown 16 hours at 37 °C for auto induction of protein expression. The cells were pelleted and resuspended with the maltose binding protein column buffer (pMAL protein expression and

purification system, NEB) and treated with lysozyme (Sigma) (250 mg/ml) for 30 min at 37 °C. The culture was then sonicated with 10 sec pulses for 2 min. The sample was centrifuged 20,000xg for 20 min and the supernatant was transferred to a clean centrifuge tube. To precipitate nucleic acids, 0.01% (w/v) linear polyethylene imine (PEI) was added to the lysate and centrifuged 15,000xg for 15 min. The lysate then was diluted to 1:6 ratio with the column buffer and purified with an amylose column as described by manufacturer (NEB). The fractions were eluted with the elution buffer (containing 10 mM maltose) and dialyzed against 20mM Tris-HCl, 50mM NaCl, and 1mM EDTA.

6.2.5.2 Chitin affinity column

The harvesting and lysing the cells were performed as described as amylose affinity column purification except that the buffer was IMPACT protein purification system column buffer. After PEI treatment, the lysate was purified with a chitin column as described by the manufacturer (NEB). Before eluting the protein fractions, the cleavage buffer was added to the column. Cleavage buffer contains 50mM DTT and incubated 24-48 hours at 4 °C. The fractions were collected with the column buffer and dialyzed against 20mM Tris-HCl, 100mM NaCl, and 1mM EDTA.

6.3 Results and Discussion

Before assaying protein binding to domain A^{ISO} RNA, we performed native polyacrylamide gel electrophoresis analysis to find the optimal folding conditions for domain A^{ISO} RNA *in vitro*.

6.3.1 *In vitro* folding conditions for domain A RNA

Domain A^{ISO} RNA appears to assume different conformations in different conditions. The *in vitro* folding of Domain A RNA depends on ionic strenght, pH,

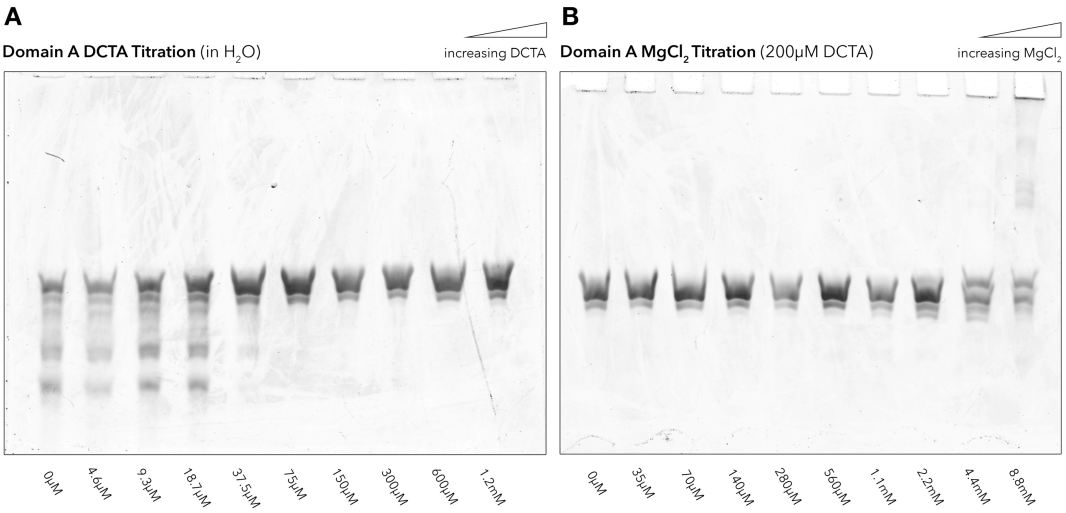

Figure 6.2 Domain A DCTA and Mg^{2+} titrations in water.
Six percent native polyacrylamide gel electrophoresis anaylsis (A) DCTA-NaOH pH 7.0 titration with increasing concentrations. Domain A^{ISO} RNA is in water (B) $MgCl_2$ titration with increasing concentrations. Domain A^{ISO} RNA is in water and 200 μM DCTA is initially added to chilate the Mg^{2+} in the beginning.

and divalent cation concentrations. Figure 6.2 shows a native polyacrylamide gel for Domain A^{ISO} RNA. The results suggest that domain A^{ISO} RNA has multiple conformations. Increasing the DCTA (1,2-Diaminocyclohexanetetraacetic acid) concentration at pH 7.0 in water leads domain A^{ISO} RNA to fold to a single conformation after 75 μM of DCTA-NaOH pH 7.0 is added. This conformational change occurs as a result of increasing Na^+ cation concentration in the solution. Monovalent cations help RNAs to form secondary structures. Figure 6.2B shows effects of divalent cations in the presence of monovalent cations. Increasing Mg^{2+} concentration causes higher order structure formation in domain A^{ISO} RNA

and this effect is seen clearly after 4mM Mg^{2+} is added to the solution as the final concentration.

When domain A^{ISO} RNA is in a buffer solution, it forms two primary

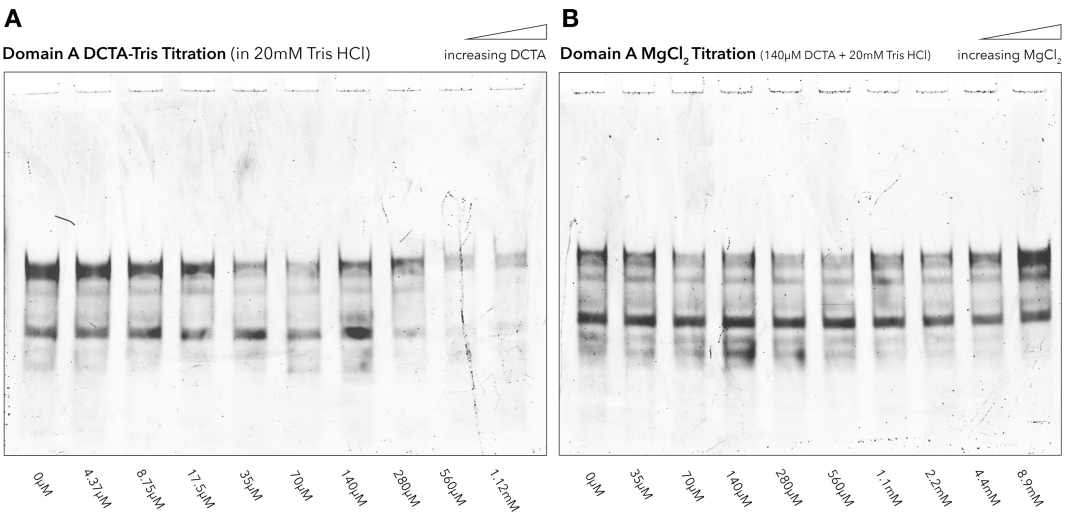

Figure 6.3 Domain A DCTA and Mg^{2+} titrations in a buffer solution.
6% native polyacrylamide gel electrophoresis anaylsis (A) DCTA-Tris pH 8.0 titration with increasing concentrations. Domain A^{ISO} RNA is in 20 mM Tris-HC pH 8.0 (B) $MgCl_2$ titration with increasing concentrations. Domain A^{ISO} RNA is in 20 mM Tris-HC pH 8.0 and 140 μM DCTA-Tris pH 8.0 is initially added to chilate the Mg^{2+} in the beginning.

conformations as seen from Figure 6.3. In 20 mM Tris-HCl pH 8.0 buffer, increasing DCTA-Tris pH 8.0 concentrations does not affect the conformation of domain A^{ISO} RNA. Since DCTA-Tris pH 8.0 does not introduce monovalent or divalent cations into the solution, increasing concentrations do not alter the folding significantly. When Mg^{2+} is added to the solution, we start seeing subtle changes in the minor bands after approximately 1mM $MgCl_2$ is added as the final concentration (Figure 6.3B). Minor bands disappear and the primary band intensity increases as a result of folding of the RNA in the presence of divalent cations.

When it comes to pH effect, low and high pHs give similar conformations for domain A^{ISO} RNA as suggested by the folding patterns of the gels. Figure 6.4 shows folding patterns for domain A^{ISO} RNA at acidic pH (5.5) and basic pH (8.0). As in the previous folding experiments, domain A^{ISO} RNA shows two major bands

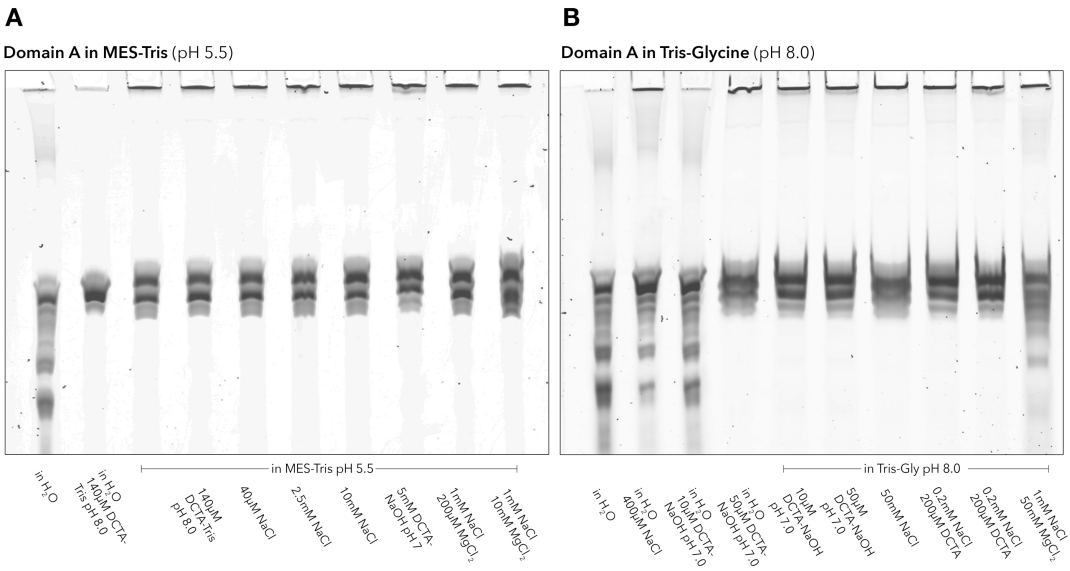

Figure 6.4 Domain A RNA folding in acidic and basic pH.
Six percent native polyacrylamide gel electrophoresis anaylsis (A) Domain A^{ISO} RNA folding in MES-Tris buffer at pH 5.5 with different monovalent and divalent conditions. (B) Domain A^{ISO} RNA folding in Tris-Gly buffer at pH 8.0 with different monovalent and divalent conditions.

in both acidic and basic pHs.

Overall, *in vitro* folding experiments of domain A^{ISO} RNA suggest that domain A^{ISO} RNA folds to a single conformation in normal monovalent and divalent concentrations such as 100-200 mM NaCl, 1-4 mM $MgCl_2$ in buffered solutions. However, domain A^{ISO} RNA in water shows many unfavorable conformations. Folding of domain A^{ISO} RNA requires pHs close to neutral pH, additionally monovalent and divalent cations.

6.3.2 *In vivo* interactions of ribosomal proteins analyzed by yeast-three hybrid system

Here we have tested the *in vivo* interactions between domain A^{ISO} and the ribosomal proteins S5 and S12 by the yeast-three hybrid assay. Briefly, the yeast-three hybrid assay is a powerful tool to monitor RNA-protein interactions *in vivo*. Target RNA, which is fused to MS2 RNA is transferred to yeast cells together with a target protein, which is fused to Gal4 activation protein. The LexA operator in *S. cerevisiae* strain YBZ-1 controls the HIS3 reporter gene. Hybrid 1 is composed of a LexA/MS2 coat protein fusion, which binds to the LexA operator. The MS2 coat protein interacts tightly to the MS2 sequence of the hybrid RNA, which is designed to carry the MS2 RNA and RNA sequence of interest, e.g., domain A^{ISO} RNA. In Hybrid 3, the yeast GAL4 transcriptional activation domain (GAD) is attached to the protein of interest (e.g., rProteins S5 or S12). The HIS3 reporter gene is expressed as a result of the *in vivo* binding of the RNA of interest to the protein of interest, which completes hybrid 2. The strenght of RNA-protein binding is evaluated by resistance to a competitive inhibitor of the HIS3 product, 3-Amino-1,2,4-triazole (3-AT), imidazoleglycerol-phosphate dehydratase.

Ribosomal proteins S5 and S12 are tertiary binding proteins as discussed earlier. They require prior bindings to other proteins to the rRNA. Our expected results are very weak or no interaction between domain A RNA and ribosomal proteins S5 and S12 *in vivo*. Our yeast-three hybrid results support this hypothesis.

6.3.2.1 Ribosomal protein S5

The binding of ribosomal protein S5 to domain A RNA was assayed *in vivo* with yeast-three hybrid system. Figure 6.5 demonstrates cell growth based on the

interaction between target protein and target RNA *in vivo*. The first bar indicates the strength of the interactions between domain A and S5 while the second bar additionally includes the inhibitor 3AT. S5-pIIIMS2 and domain A-pACT2 are negative control experiments. In those cases, plasmid pIIIMS2 does not include

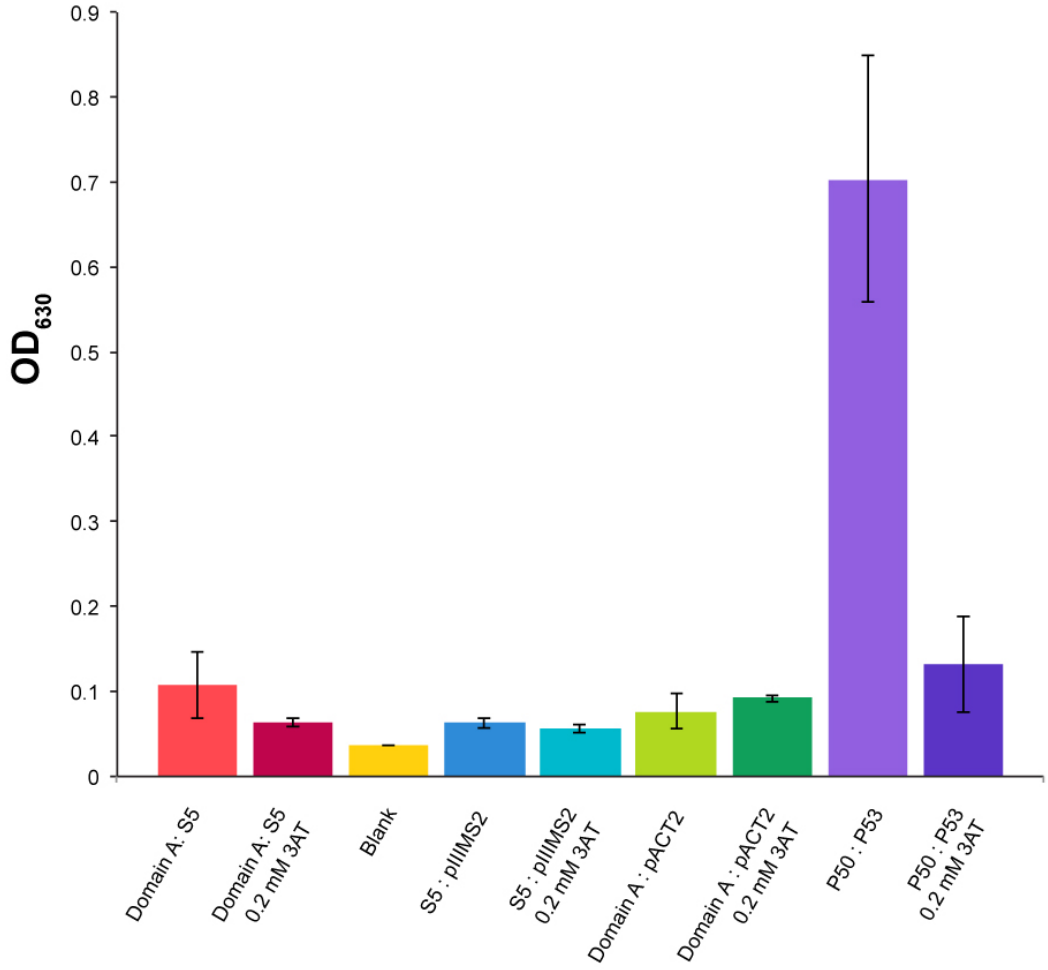

Figure 6.5 Yeast-three hybrid interactions between domain A and ribosomal protein S5.
(From left to right) The first lane shows cell growth indicating the interaction between domain A and S5 while the second lane shows the same interaction in the presence of inhibitor 3AT. The third lane is blank. The forth and fifth lane are negative control for protein S5 and the later includes inhibitor 3AT. The sixth and seventh lane are negative control for the RNA without and with the inhibitor 3AT respectively. The eighth and nineth lane are positive control without and with the inhibitor 3AT. All absorbance reads are at 72 hours incubation.

is growing media only. The bar showing p50-p53 is the positive control reaction demonstrating the strong binding between RNA and protein. All results are absorbance reading of the corresponding yeast cells at 630 nm after 72 hours of incubation.

Compared with the positive control reaction, domain A and ribosomal protein S5 shows a very minimal signal, indicating very weak or no binding. Since ribosomal protein S5 is a tertiary binder, these results support the hypothesis that specific binding of ribosomal protein S5 to domain A RNA requires primary and secondary binding proteins to be bound to 16S rRNA first.

6.3.2.2 Ribosomal protein S12

Ribosomal protein S12 binding to domain A RNA is assayed *in vivo* with yeast-three hybrid system. Figure 6.6 demonstrates the extent of interaction between target protein and target RNA *in vivo* as indicated by cell growth. As with domain A and S5, the first bar indicates the extent of interaction between domain A and S12 while the second bar includes the inhibitor 3AT. The bars showing S12-pIIIMS2 and domain A-pACT2 the extent of interactions are negative control experiments. In those cases, plasmid pIIIMS2 does not include the target RNA and plasmid pACT2 does not include the target protein. The blank lane is growing media only. The bar showing p50-p53 is the positive control reaction demonstrating the strong binding between RNA and protein. All results are absorbance reading of the corresponding yeast cells at 630 nm after 72 hours of incubation.

Compared with the positive control reaction, domain A and ribosomal protein S12 show very minimal signal indicating no binding *in vivo*. The ribosomal protein S12 is also a tertiary binding protein. Specific binding of ribosomal protein S12 to domain A RNA requires the primary and secondary binding proteins to be

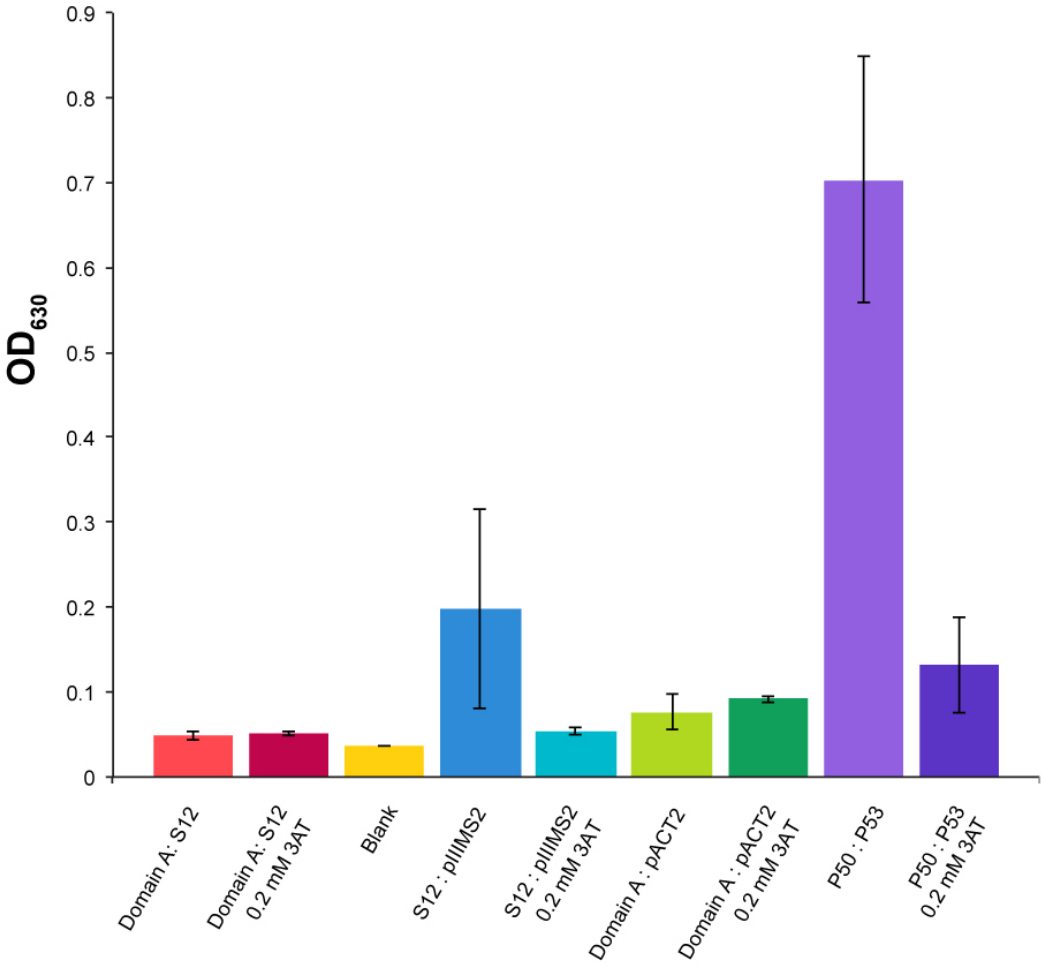

Figure 6.6 Yeast-three hybrid interactions between domain A and ribosomal protein S12.
(From left to right) The first lane shows the cell growth indicating the interaction between domain A and S12 while the second lane shows same interaction in the presence of inhibitor 3AT. The third lane is blank. The forth and fifth lane are negative control for protein S12 and the later includes inhibitor 3AT. The sixth and seventh lanes are negative control for the RNA without and with the inhibitor 3AT respectively. The eighth and nineth lane are positive control without and with the inhibitor 3AT. All absorbance reads are at 72 hours incubation.

bound to 16S rRNA first.

When it comes to comparison between S5 and S12, both proteins show very weak interactions with domain A RNA. S5 gives slightly a bigger signal than S12 however comparison with negative control experiments reveals that the difference between S5 and S12 is not significant.

6.3.3 *In vitro* interactions of ribosomal proteins analyzed by electrophoretic mobility shift assay

6.3.3.1 Ribosomal protein S5

The ribosomal protein S5 is composed of 162 amino acids with a theoretical isoelectronic point (pI) of 10.05. The number of positively charged residues is 24, which 16 are arginine and 8 are lysine residues. This indicates S5 is favorable for nucleic acid binding. We hypostasize that, since the positively charged residues are highly represented in the protein, interactions between S5 and RNA are highly probable and this leads unspecific binding interactions to be favorable.

Here we have tested the binding of S5 to domain A RNA *in vitro* by electrophoretic mobility shift assay (EMSA). Figure 6.7 demonstrates the interactions between domain A RNA and ribosomal protein S5 (maltose binding protein fusion). The first lane out the left is the S5 fusion protein alone. The second lane is domain A RNA alone. The third through seventh lanes are increasing stoichiometric ratio between domain A and S5 fusion, keeping the amount of RNA constant. The last lane is the control reaction showing lack of domain A and

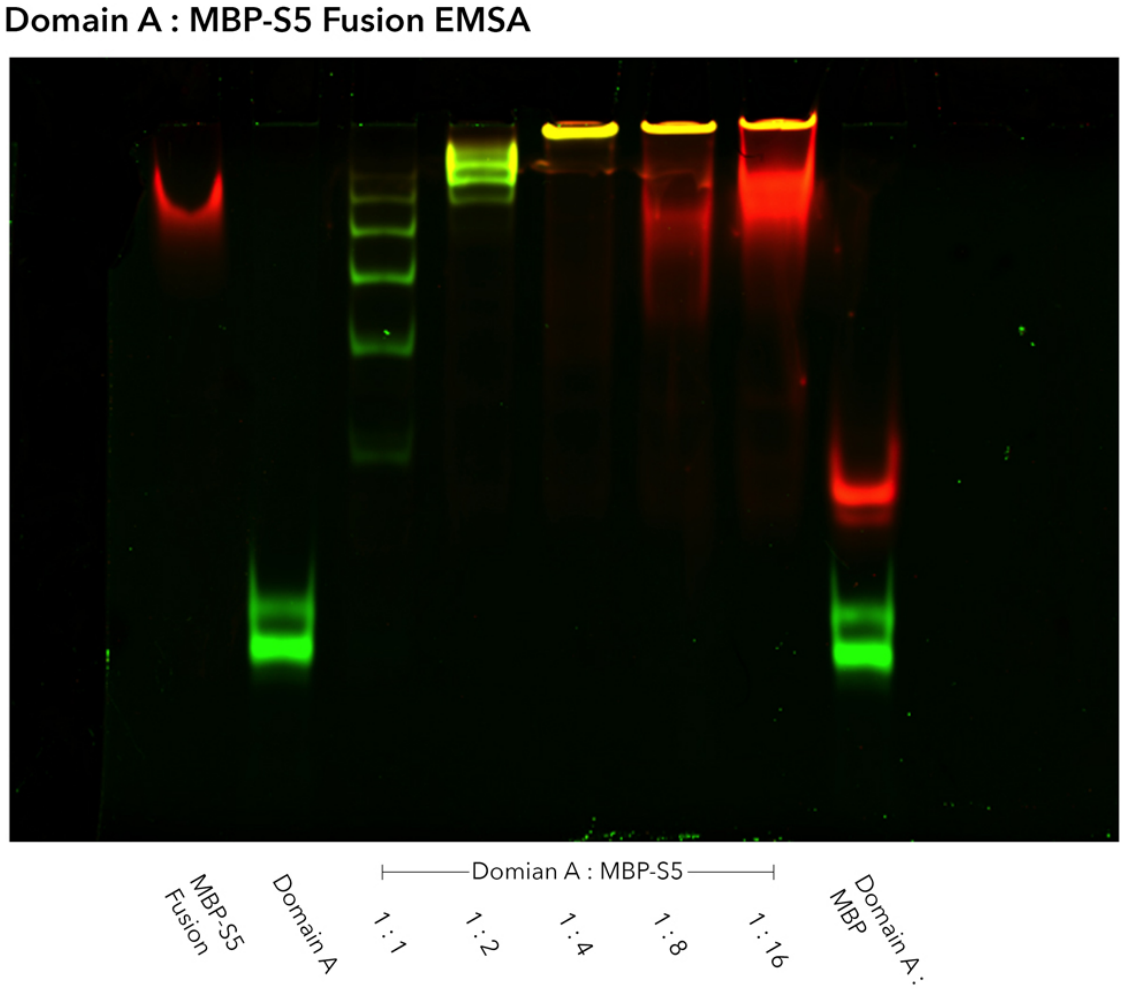

Figure 6.7 Electrophoretic mobility shift assay for domain A^{ISO} RNA and S5 fusion protein.
From left to the right first band shows S5 fusion alone. The second band is domain A^{ISO} RNA alone. The third through eighth band show an increasing concentration of the protein while the RNA is constant. The last band shows a negative control for maltose binding protein and domain A^{ISO} RNA.

maltose binding protein interactions. The "ladder pattern" of 1 to 1 ratio shows that there are multiple proteins binding to one RNA molecule. To be more specific, counting from the bottom band, 1 to 1 through 1 to 6 RNA to protein-binding ratio is seen in the third lane. When the protein concentration increases dramatically, the ladder pattern disappears in that the RNA becomes fully saturated with the protein. This phenomenon is shown previously for such proteins

that bind to nucleic acids unspecifically (Bendak et al., 2012). The control reaction in the last lane reveals no interaction between domain A RNA and maltose binding protein. This is an indication of a domain A – S5 complex without the maltose binding protein interference.

Overall, we hypothesize that ribosomal protein S5 binds unspecifically to domain A RNA *in vitro*. These results support our hypothesis of the positive charge effect.

6.3.3.1.1 P4P6 RNA and 12-mer duplex RNA interactions

To further investigate and confirm the unspecific binding of ribosomal protein S5, we have tested of different RNAs such as P4-P6 domain of group I intron and a 12-mer duplex RNA with S5. P4-P6 RNA is 160 nucleotides long, which is in a similar range of 133 nucleotide long domain A RNA while the 12-mer duplex RNA represents a shorter RNA length that S5 binds to. Figure 6.8 shows *in vitro* bandshift between S5 fusion protein and P4-P6 RNA or 12-mer duplex RNA. In the case of P4-P6 RNA, we have observed a very similar pattern as we have seen in domain A –S5 interactions. Multiple S5 proteins bind to one P4-P6 RNA molecule (Figure 6.8A). This result supports the hypothesis that ribosomal protein S5 binds to domain A RNA unspecifically *in vitro*. Furthermore, we have tested the interaction of S5 with a small RNA; a 12 base paired double stranded duplex. Figure 6.8B shows the bandshift between RNA duplex and S5 fusion protein. In this native gel image, the RNA is not stained due to its length so that is not visible in the gel. However, protein behavior reveals a similar effect but with fewer bands. The 1 to 1 ratio changes to a more complex binding pattern extending from 2

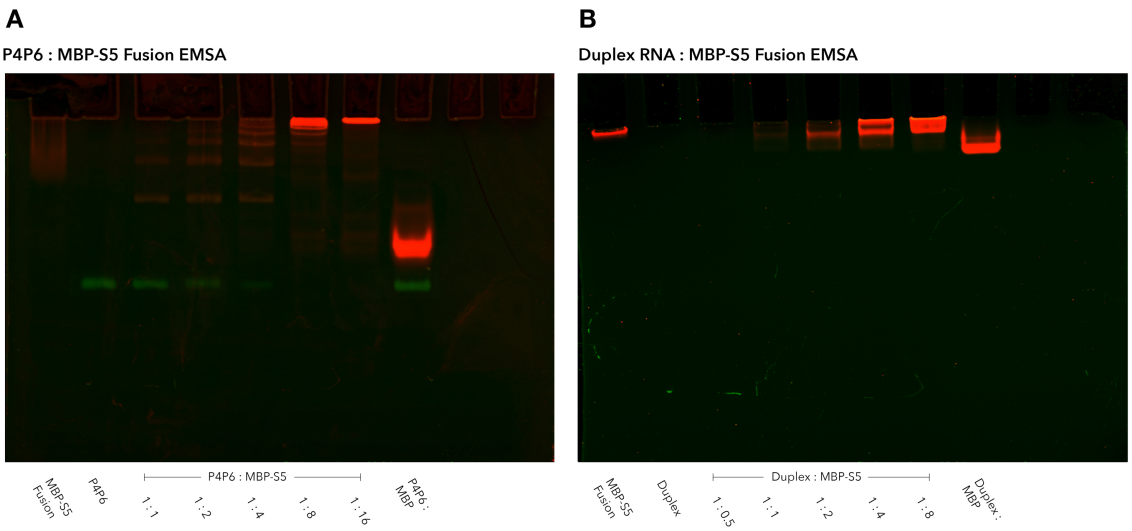

Figure 6.8 Electrophoretic mobility shift assay for P4-P6 RNA / 12-mer duplex RNA and S5 fusion protein.
(A) Interactions between P4-P6 domain of group I intron RNA and S5 fusion protein. From left to right, the first band shows S5 fusion alone. The second band is P4-P6 RNA alone. The third through eighth band show increasing concentration of protein while the RNA is constant. The last band shows negative control for maltose binding protein and P4-P6 RNA. (B) Interactions between 12-mer duplex RNA and S5 fusion protein. From left to right, the first band shows S5 fusion alone. The second band is 12-mer duplex RNA alone. The third through eighth band show increasing concentration of protein while the RNA is constant. The last band shows negative control for maltose binding protein and 12-mer duplex RNA. Because of the size and amount of RNA loaded into gel, the RNA staining is not visible. However, protein behaviour reveals the unspecific binding between S5 fusion and 12 mer duplex RNA.

bands in 1 to 4 ratio between duplex RNA and S5 fusion protein respectively. These results suggest that the unspecific binding of S5 to RNAs is increasing with the increasing length of the RNAs as a result of available surface area. It has ben previously reported that ribosomal protein S5 shows certain binding properties to RNAs (Bycroft, Grunert, Murzin, Proctor, & Johnston, 1995; Fukushi et al., 2001; Ramakrishnan & White, 1992).

Whereas the interaction between domain A^{ISO} and S5 *in vivo* was very weak, the unspecific binding potential of S5 to nucleic acids might cause a

competition between domain A RNA and cellular nucleic acids in yeast. As a result of this unwanted interactions, the yeast-three hybrid signal was not high enough. This hypothesis is also supported by the negative control reaction in yeast-three hybrid system when protein S5 and especially S12 were transferred and expressed in yeast in the absence of domain A^{ISO} RNA (Figure 6.5 and Figure 6.6). The negative control signal was the same as the normal reaction signal indicating unspecific binding of those proteins.

6.3.3.1.2 SHAPE reaction with S5 fusion protein

To elaborate the interactions between S5 fusion protein and domain A RNA, we have saturated the RNA with the protein and performed SHAPE reaction for the domain A^{ISO} RNA. Figure 6.9 clearly reveals that the S5 binding to domain A^{ISO} RNA elevates the SHAPE reactivity of the nucleotides in the central pseudoknot dramatically. Since other secondary structures of domain A^{ISO} were not affected from the protein binding, the SHAPE results suggest that S5 binding melts the central pseudoknot. In other words, tertiary folding of domain A^{ISO} RNA is disrupted by saturation with the ribosomal protein S5. We have discussed the significance of the central pseudoknot and its folding in earlier chapters. These results suggest that positively charged residues of ribosomal protein S5 affect the tertiary folding of the central pseudoknot when the protein concentration is high. Figure 6.9C shows the SHAPE reactivity for each nucleotide. The central pseudoknot nucleotides are demonstrated with a thin line and it is clear that the reactivity in those nucleotides after the S5 saturation is elevated drastically indicating melting in the tertiary structure of the pseudoknot.

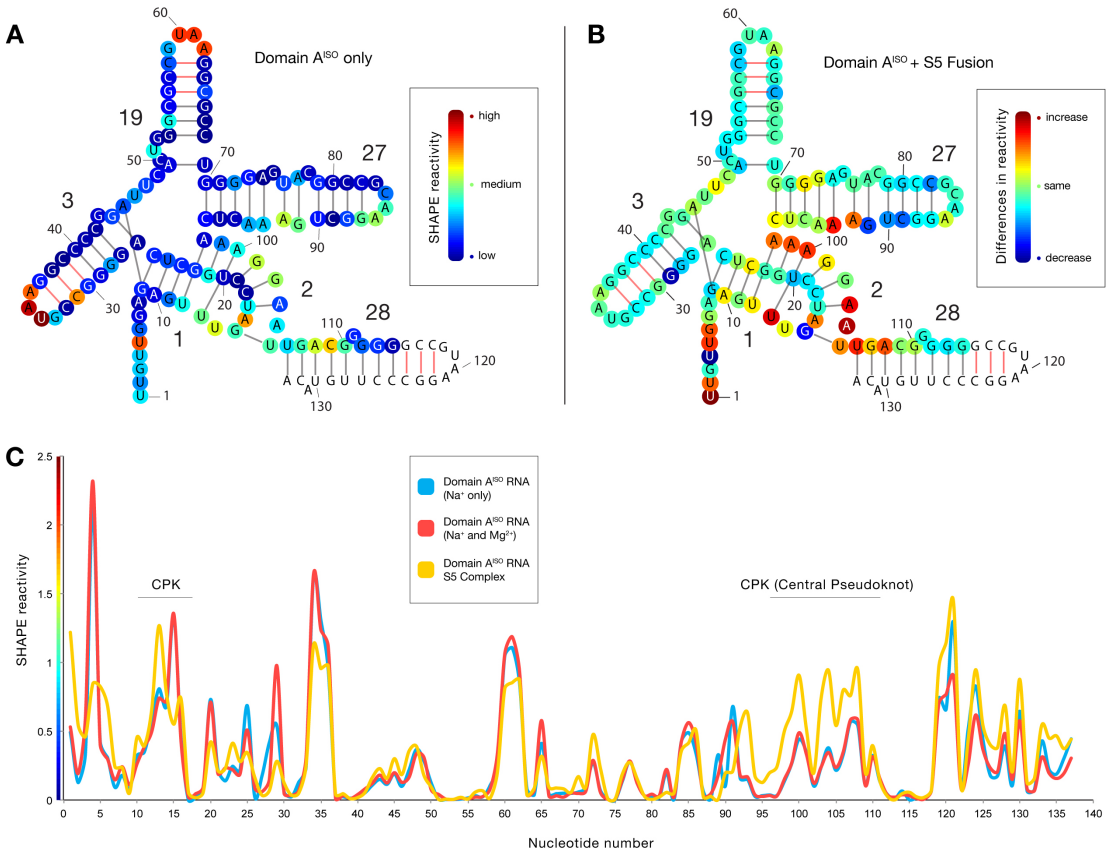

Figure 6.9 SHAPE reactivity of the intact domain A^{ISO} and domain A^{ISO} ribosomal protein S5 fusion complex.

Base pairs predicted from the secondary structure of the intact SSU are indicated by black lines. Presumed base pairs in the linkers are indicated by red lines. Helix and nucleotide numbers are indicated. (A) Reactivity of domain A^{ISO} in the presence of Na^+ and Mg^{2+}. The red circles indicate high reactivity while the blue circles indicate low reactivity. The color scale is shown in the outbox. (B) Difference in SHAPE reactivity upon addition of S5 fusion protein in saturated 1 to 8 RNA to protein ratio. Red indicates an increase in reactivity, while blue indicates a decrease. Green indicates no change. The coloring scheme is shown in the outbox. Data for domain A^{ISO} – S5 fusion complex are acquired in the presence of both Na^+ and Mg^{2+}. Data were not accessible for the uncolored nucleotides. The primer binding tail is omitted for clarity. (C) Processed SHAPE reactivity of domain AI^{SO} RNA with sodium only (blue), sodium and magnesium (red), and S5 fusion protein (yellow). Corresponding SHAPE reactivity color scale is shown on the y-axis. The central pseudoknot nucleotides are indicated with the thin black lines.

6.3.3.1.3 Circular dichroism of domain A^{ISO} with S5 fusion protein

Further evidence for disruption of the tertiary folding came from CD experiments. We have scanned the S5 fusion saturated domain A^{ISO} RNA in the presence of 1mM Mg^{2+}. Figure 6.10A illustrates the change in CD signal after addition of S5 fusion protein to domain A^{ISO} RNA. Diagnostic signal at 265 nm for the RNA – protein complex (blue) decreases approximately 50 % of domain A^{ISO} RNA alone (red). A similar effect was previously observed once we applied mutations to disrupt the central pseudoknot of domain A^{ISO} RNA (Figure 6.10B). Considering the SHAPE reactivity change in this region and comparing the same effects of the C18A and helix 28 truncation mutations, 50 % drop in CD signal is an indication of the melting of the central pseudoknot upon binding of ribosomal protein S5.

6.3.3.2 Ribosomal protein S12

On the other hand, ribosomal protein S12 is composed of 135 amino acids and its theoretical pI was calculated 11.01. There are 35 positively charged amino acid residues in the S12 structure (14 arginine, 21 lysine residues). Similar charge properties of S12 give similar results with S5 on EMSA experiments (Data not shown).

6.4 Conclusion

Domain A of 16S rRNA is located in the core of SSU. Out of 21 ribosomal proteins in bacterial species, S5 and S12 bind to domain A in 30S ribosomal subunit. It has been shown that S5 and S12 are tertiary binding proteins meaning that they require other proteins bind to 16S rRNA first. Here we have tested the

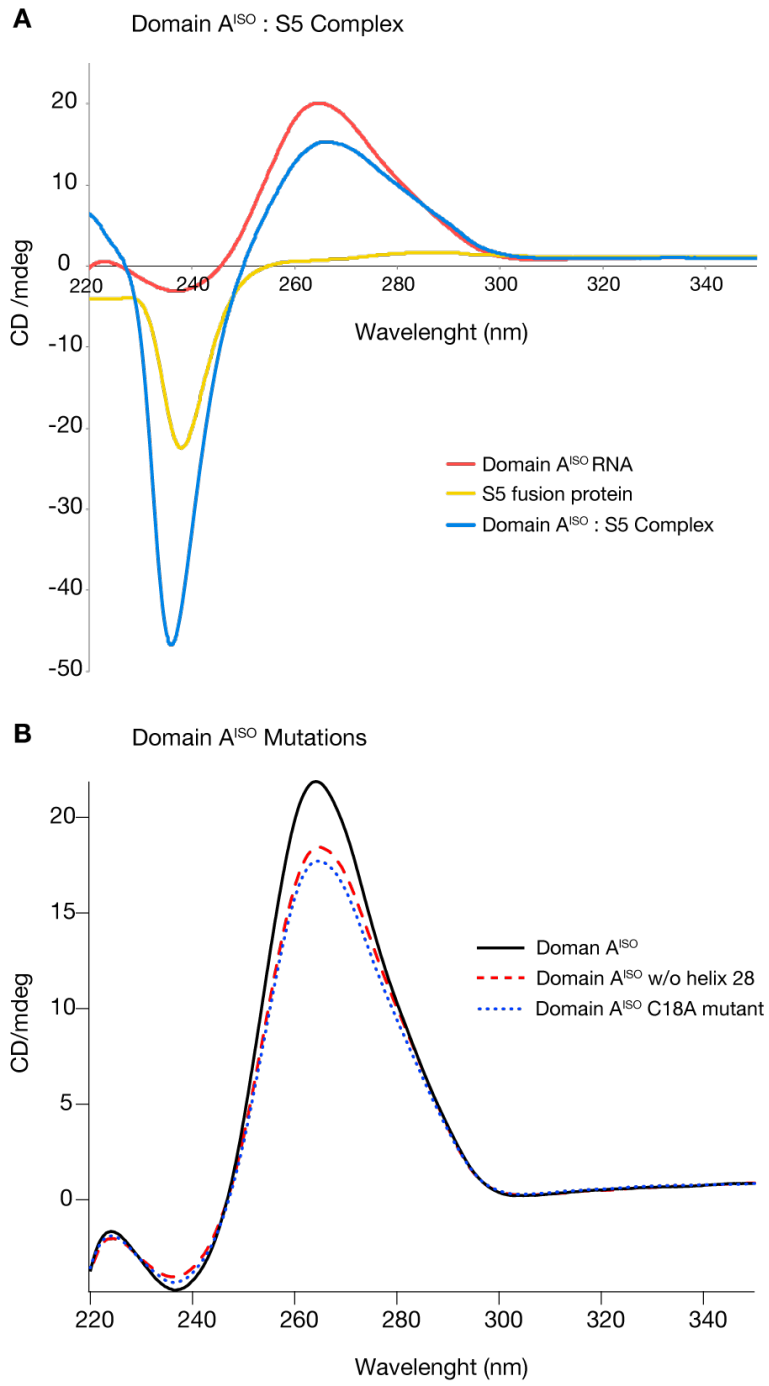

Figure 6.10 Circular dichroism spectroscopy of domain A^{ISO} RNA – ribosomal protein S5 complex.
(A) CD spectra of domain A^{ISO} RNA alone (red), ribosomal protein S5 fusion alone (yellow), and domain A^{ISO} RNA – S5 fusion complex (blue). All CD scans performed in the presence of 1 mM Mg^{2+} (B) CD spectra of domain A^{ISO} RNA (solid black), the C18A mutant (dotted blue), and helix 28 truncated (dashed red) domain A^{ISO} RNAs in the presence of 1 mM Mg^{2+}.

binding of these proteins to domain A^{ISO} RNA *in vivo* and *in vitro*. Our yeast-three hybrid results suggest that S5 and S12 are not interacting with domain A *in vivo*. A possible explanation of this is the positive charge potential of those proteins. In other words, S5 and S12 have nucleic acid binding properties based on their high positive charge. This leads to an unspecific binding of those proteins to any RNA. In order to elaborate this hypothesis *in vitro*, we have performed EMSA, SHAPE, and CD experiments for S5 as a model protein. Our EMSA results show that S5 fusion protein binds to domain A^{ISO} RNA unspecifically. Increasing concentration of the protein while the RNA is constant reveals multiple proteins binding to one RNA molecule. We have also tested different RNAs such as P4-P6 domain of group I intron and a 12-mer duplex RNA. P4-P6 RNA showed a similar multiple binding pattern as domain A^{ISO} RNA confirming the unspecific binding. On the other hand, 12-mer duplex RNA and S5 complex showed that 1 and maximum 2 proteins bind to one RNA molecule. This results support the idea that S5 has a binding affinity to RNA. Furthermore, based on the surface area of the RNA molecule, multiple proteins can bind to one RNA molecule. Binding motifs of RNA binding proteins and unspecific binding of S5 to other RNAs are shown previously (Bycroft et al., 1995; Draper & Reynaldo, 1999; Fukushi et al., 2001; Maris, Dominguez, & Allain, 2005; Soding & Lupas, 2003).

To further investigate the binding reaction between domain A^{ISO} RNA and ribosomal protein S5, we have performed SHAPE and CD experiments on protein saturated domain A^{ISO} –S5 fusion complex. Our results suggest that the binding of S5 fusion protein melts the central pseudoknot. We hypothesize that positive charged residues of S5 are breaking up the tertiary structure of the central

pseudoknot. Since the central pseudoknot is not very stable due to its flexibility (even one single mutation can disrupt its structure), high positive charges can cause ablation of the structure. In other words, folding of the central pseudoknot is affected by positively charged ribosomal protein S5. This hypothesis also explains why the ribosomal proteins S5 and also the similarly charged S12 are not primary binding proteins in 30S assembly (Bunner et al., 2010; Mulder et al., 2010). Otherwise, primary binding of these proteins would retard the 16S rRNA folding by destabilizing the central pseudoknot in 30S assembly in that domain A acts as a hub where other domains radiate from it. This hypothesis is also consistent with the fact that domain A is buried in the core of 16S rRNA requiring other proteins to bind and stabilize the overall 16S structure prior to binding of the highly basic proteins S5 and S12.

CHAPTER 7

EVOLUTIONARY MODEL OF THE CENTRAL PSEUDOKNOT AND SMALL RIBOSOMAL SUBUNIT

There is no doubt that the evolution of ribosome is one of the most significant events in biology. Here we propose an evolutionary model for the SSU rRNA following the Petrov and coworkers LSU evolution model (Petrov et al., 2014). Our model is based on the insertion fingerprints, which are observed by eukaryotic expansions. Moreover, our model explains the evolution of the central pseudoknot and subunit association.

7.1 Introduction

7.1.1 Ribosomal Growth

The ribosome is life's central macromolecule and keeps its own history written in its core. Many researchers suggested that rRNA grows and this growth is occurring by accretion. In other words, there are pieces added to rRNA while keeping the core structure unchanged. Bachellerie (Michot, Qu, & Bachellerie, 1990), Steinberg (Bokov & Steinberg, 2009) and others proposed a similar hypothesis which states that the LSU RNA is increasing in size by accretion. One major application of this process is the possibility of tracking the ribosome structure through past events to reveal clues about ribosomal origins.

Here we showed similar accretion results for the SSU RNA. For instance, *Saccharomyces cerevisiae* has six SSU rRNA expansion regions, more specifically six expansion segments (es3, es6, es7, es9, es10, es12) (Gerbi, 1996; Hassouna,

Michot, & Bachellerie, 1984; Melnikov et al., 2012). These expansion segments are extended or branched from common core helices. In some cases, expansion segments are directly branched from a helix, without perturbing the helical structure of the trunk helix, as an insertion (es3a, es3b, es6a, es6b, es6c, es6d, and es10). Other cases occur as an extension of the core helix (es7, es9, and es12).

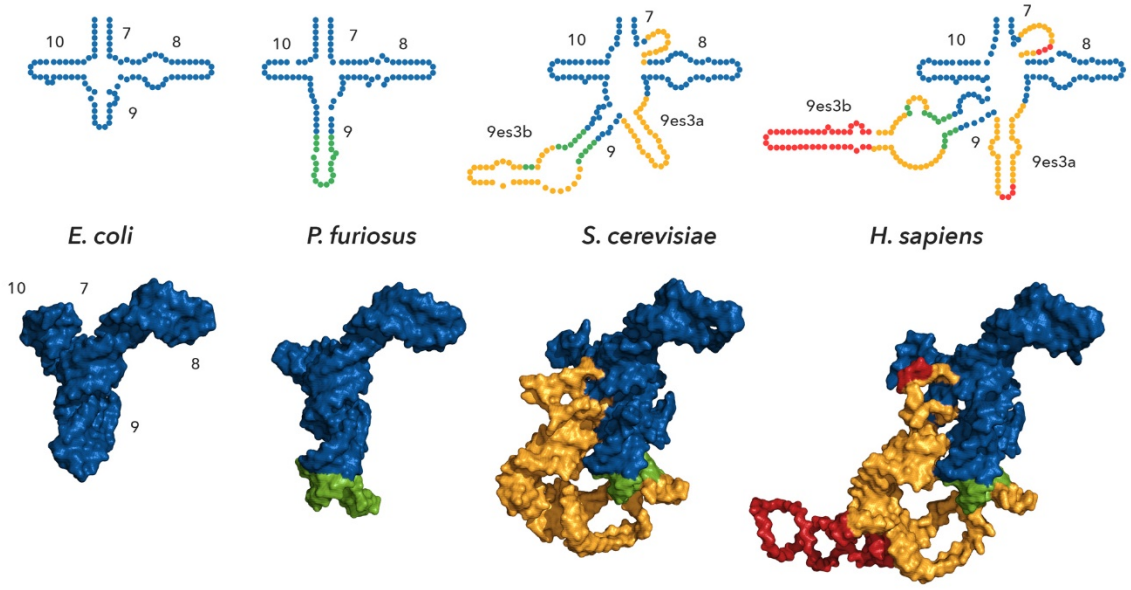

Figure 7.1 Eukaryotic expansion segment es3 and its growth.
Es3 grows from common core helix, in this case, helix 9. From *E.coli* to *P.furiosus*, expansion of the helix is green in both 2D and 3D structures. When the complexity of the species increases, we see a growth on the same helix as eukaryotic expansion segment 3 (yellow). From *S.cerevisiae* to *H.sapiens*, we can see additional growth shown with red.

Recursive expansions are added to the core of the rRNA that gradually enlarge the size of the rRNA. Once expansion occurs, the underlying core RNA stays unaltered. Figure 7.1 demonstrates the expansion by accretion using expansion segment es3 as an example. This superimposition and comparison of core rRNA segments from different species can be used to infer structures of

common ancestors (Fitch, 1971). Petrov and coworkers previously used the assumption that the structure of the rRNA of an ancestor can be approximated by structural elements common to the progeny (Petrov et al., 2014). Using this assumption, it is possible to create an evolutionary timeline for rRNA expansions by employing three-dimensional structures. Since the rRNAs are highly complex and branched, it is difficult to interpret the homologous expansions.

7.1.2 Eukaryotic Expansions

Experimental data for the evolution pathway of rRNA is already provided naturally by eukaryotic expansions. An example of this pathway is shown in Figure 7.1 focusing on common core helices 7, 8, 9, and 10. In *E. coli*, this example region of the 16S rRNA elements is composed of 92 nucleotides. It is a good representation of the common core with the exception of slight changes in helix 10 (Selmer et al., 2006). Helices 7 through 10 are expanded by extension of helix 9 from the bacteria *E.coli* to the archean *P. furiosus*, adding 11 nucleotides (Armache et al., 2013). Furthermore, from archea to eukarya we observe the additional expansion of helix 9 through the addition of three more helices, which become the eukaryotic expansion segment es3. With the addition of es3 to the common core (helices 7 through 10), this rRNA segment become174 nucleotides in *S. cerevisiae* (Ben-Shem, de Loubresse, et al., 2011), and this size is mostly maintained in other lower eukaryotes. Moreover, in higher eukaryotes such as mammals, es3 grows further and adds additional fragments (es3a and es3b), reaching 223 nucleotides in *H. sapiens* (Anger et al., 2013). From *E. coil* to *H.sapiens* this expansion example illustrates rRNA growth by accretion on an increasing frozen core (Figure 7.1). Once the addition occurs by accretion,

common core rRNA become frozen and unchanged. A representation of this example is shown in 2D and 3D in Figure 7.1.

On the other hand, increasing size of the expansion segments does not follow the same phenomenon although it is observed universally for most cases. In other words, rRNA does not always follow localized monotonic growth in size. For example, es6 of *D. melanogaster* and *T. brucei* contain an additional helix (21es6b1) as we define previously (Petrov et al., 2014). This helix expansion is absent in the mammalian SSU RNA.

The common core of the rRNA is approximated from the *E.coli* for both LSU and SSU RNAs (Melnikov et al., 2012; Michot et al., 1990). The common core rRNA is conserved in sequence, secondary structure (Mears et al., 2002) and three-dimensional structure (Hsiao, Mohan, Kalahar, & Williams, 2009) over the whole phylogenetic tree of life. Unlike the common core rRNA, the size of the ribosome is increasing through species starting from bacteria and archaea to lower and higher eukarya. Growth in the LSU and SSU is not the same. Even if the process is similar for both, growth of SSU RNA in size is remarkably less than growth of LSU RNA size. When it comes to the regional identification of the expansion segments, insertions mainly occur in the surface regions of the ribosomal subunits (Gerbi, 1996; Hassouna, Michot, & Bachellerie, 1984; Melnikov et al., 2012).

In SSU RNA, there are six eukaryotic expansion segments (es3, es6, es7, es9, es10, and es12) in *S. cerevisiae* compared to *E. coli*. From those expansion elements, es7, es9, and es12 are examples of growth from existing helices of the common core rRNA while es3a, es3b, es6a, es6b, es6c, es6d, and es10 are

examples of branching from the helices of the common core rRNA. These observations were made by superimposition of the 3D structures.

7.1.3 Ancestral Expansion Segments

Previously, Petrov and coworkers investigated the common core of LSU rRNA for expansion elements, which are called insertion fingerprints (Petrov et al., 2014). It has been suggested that the common core of the ribosomal RNA was completed around approximately 4 billion years ago (Clark, Tague, Ware, & Gerbi, 1984; Hassouna et al., 1984; C. Hsiao et al., 2009; Melnikov et al., 2012). The common core RNA occurs in all cytoplasmic ribosomes. Its primary sequence but more significantly its secondary structure (Hsiao et al., 2009; Mears et al., 2002) and also its three-dimensional structure (Hsiao et al., 2009; Melnikov et al., 2012) is highly conserved through all species. Petrov and coworkers reported previously an evolutionary pathway of LSU common core rRNA by insertion fingerprints tracing backwards in time the formation of the peptidyl transferase center, the catalytic heart of the ribosome (Petrov et al., 2014).

Ancestral expansion segments represent the evolutionary pathway of ribosomal growth by accretion. The ancestral expansion segments are designed based on the eukaryotic expansions. As illustrated in Figure 7.1, eukaryotic expansions have a unique insertion fingerprint. Insertion of the new segment does not perturb the existing common core RNA. By using the same phenomenon, Petrov and coworkers traced the evolution of the ribosome by insertion fingerprints for LSU common core rRNA (A. S. Petrov, C. R. Bernier, C. Hsiao, et al., 2014). Here we have applied the same process in SSU common core rRNA by tracing the ancestral expansion segments and putting them into a

temporal order. Starting from the decoding center of the SSU rRNA, ancestral expansion segments reveal the pattern of growing in the SSU common core

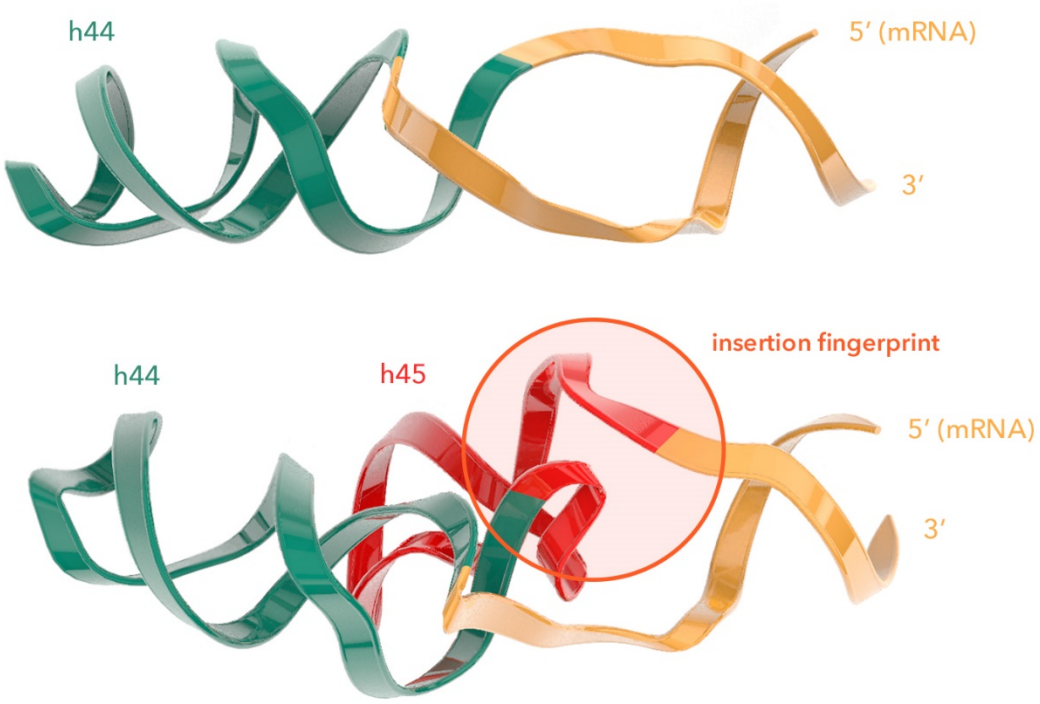

Figure 7.2 Insertion fingerprints for SSU rRNA growth.
Ancestral expansion segment 1 is shown on top which is composed of helix 44 (teal) and the putative helix (yellow). Ancestral expansion segment 2 is seen on bottom after insertion of helix 45. The insertion site is circled.

rRNA. An example insertion fingerprint is illustrated in Figure 7.2. Helix 45 of SSU rRNA of *E. coli* is inserted between the 3' end of 16S rRNA and helix 44 without perturbing the common core rRNA helix. These insertion fingerprints are very similar to their LSU counterparts. By comparing nature's eukaryotic expansion segments and how they branched out, we can infer that ancestral expansion segments followed the same process, one that led the ribosome to grow by accretion.

7.2 Materials and Methods

7.2.1 Secondary structures

All secondary structures of LSU and SSU rRNAs used in this study are taken from our public gallery (http://apollo.chemistry.gatech.edu/RibosomeGallery/) and all data are mapped on our in-house RiboVision server (Bernier et al., 2014; Petrov et al., 2014; Petrov et al., 2013).

7.2.2 Three-dimensional structures

Three dimensional structures of ribosomes were obtained from the PDB database (PDB IDs 1JJ2 (Ban, Nissen, Hansen, Moore, & Steitz, 2000), 4V6U (Armache et al., 2013), 4V9D (Dunkle et al., 2011b), 4V88 (25), 4V6W, 4V6X (Anger et al., 2013). Global and local superimpositions were performed using the built-in cealign functionality of PyMOL with default settings (Schrodinger, 2010).

7.3 Results and Discussions

We have proposed an evolutionary pathway for SSU rRNA by using ancestral expansion segments and insertion fingerprints. Our model explains independent evolution of SSU and LSU through early phases. We have described the completion of SSU evolution in 6 phases for the common core rRNA. Our results are supported by the integrity and similarity of eukaryotic expansion segments.

7.3.1 Evolutionary Model for the Small Ribosomal Subunit

7.3.1.1 Ancestral expansion segments

Here we propose an evolutionary model and pathway for the SSU rRNA. As a common core for all SSU rRNAs, we have evaluated the pathway from *E.coli*, which is an approximation of the first complete ribosome. After identifying the insertion fingerprints, we designated the ancestral expansion segments starting from the functional core of the SSU rRNA, the decoding center. The decoding center is located in the upper stem of helix 44 and is responsible for mRNA decoding during protein synthesis.

By considering the differences between cytosolic and mitochondrial ribosomes, helix 44 varies in length. This fact is reflected in our evolutionary model as division of ancestral expansion segment 1 into three different insertion elements belonging to three different phases.

Figure 7.3 shows the whole evolutionary model for the SSU rRNA through ancestral expansion segments. According to our model, ancestral segment 1 expanded with the ancestral expansion segment 2, which is helix 45 of SSU rRNA. Helix 2 and helix 28, as we elaborated in earlier chapters, form continuous stacking interactions, and inserted as ancestral expansion segment 3. Further expansion came from helix 27, which forms the ancestral expansion segment 3a. Helix 19 and helix 3 are inserted as ancestral expansion segment 4. Helix 25 is inserted as ancestral expansion segment 4a. Completion of the first three phases of the SSU evolution was achieved by the insertion of helix 24 as ancestral expansion segment 5. Previously, Noller proposed a pattern of continuity of

stacking for the SSU rRNA and our SSU decomposition is generally consistent with this pattern (Noller, 2005).

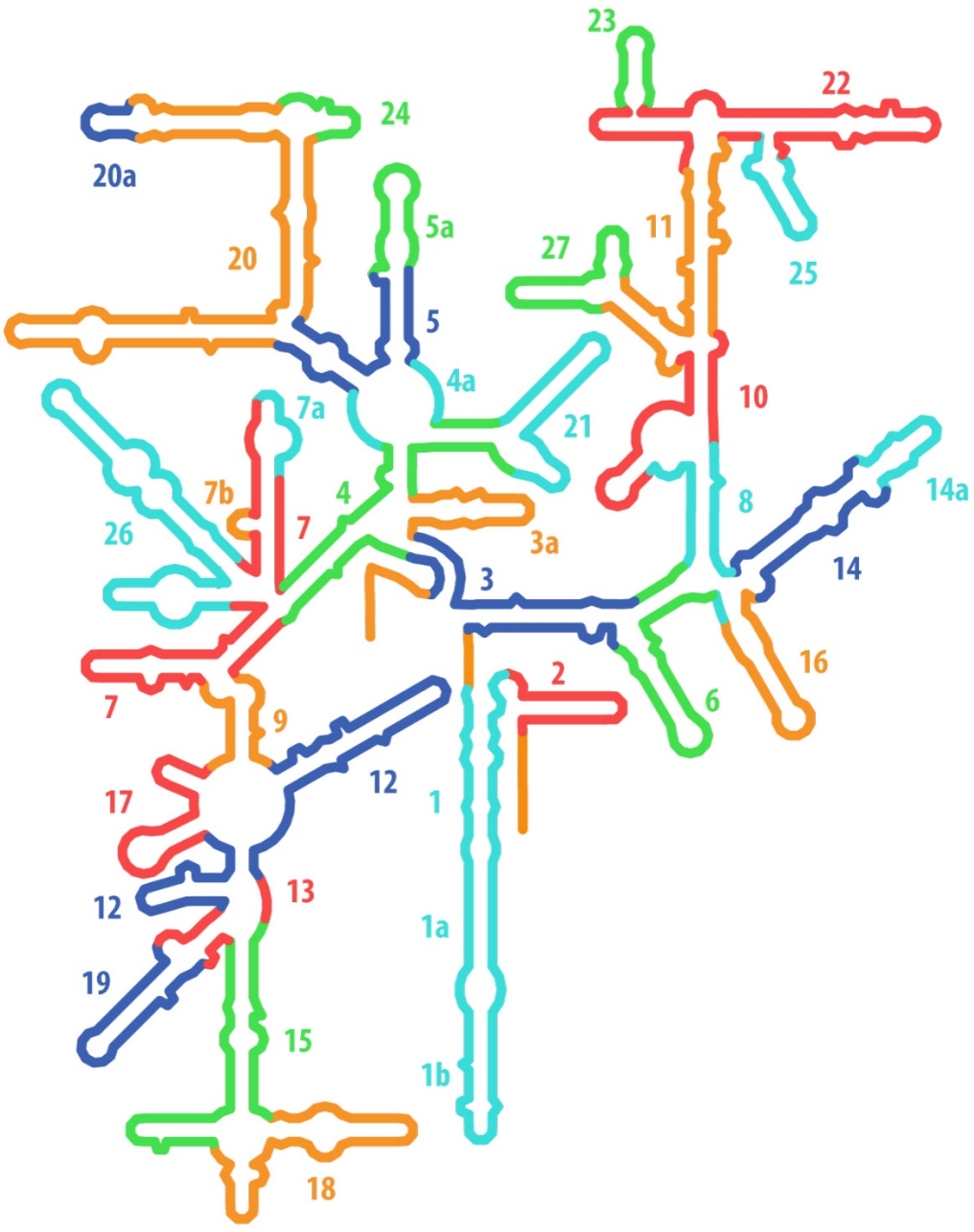

Figure 7.3 Ancestral expansion segments of SSU rRNA and their temporal order. Ancestral expansion segments 1 to 27 with 5 different colors to distinguish the transition between segments. Secondary structure is represented as countur lines and taken from the in-house Ribovision server.

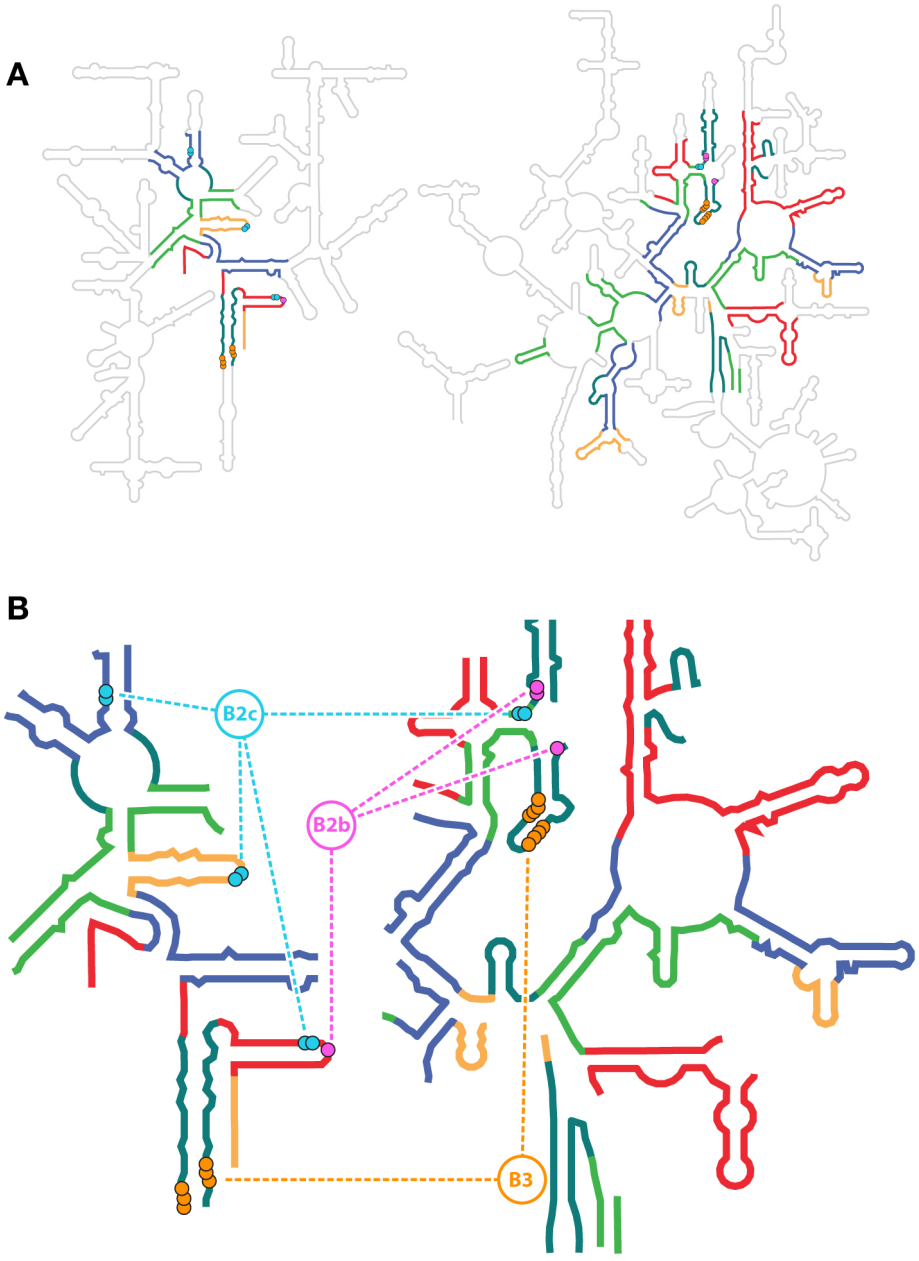

Figure 7.4 Ancestral intersubunit bridges B3, B2b, and B2c.
A) Overview of the ancestral intersubunit bridges. RNA segments correspond to phase 3 of our SSU and LSU evolution model are represented as colors while the rest of the secondary structure is grey. B) Close view of phase 3 segments showing the contact points for the ancestral intersubunit bridges.

7.3.1.2 Phases of SSU evolution

In our evolutionary model, we have divided the process of ancestral expansions into 6 different phases following the previous work of Petrov and coworkers for LSU rRNA (Petrov et al., 2014).

Figure 7.5 demonstrates the phases of SSU evolution. Unlike the LSU evolution phases, the first 3 phases of the SSU evolution model is shown as a combined phase 1 through 3. We suggest that the coevolution of the SSU and LSU occurred up through the phase 3. In other words, at the end of the phase 3 of ribosomal evolution, the SSU and LSU rRNAs were capable of forming the ribosome assembly via ancestral intersubunit bridges.

Ancestral intersubunit bridges B3, B2b, and B2c are the first interaction elements between the ribosomal subunits in early evolution. Figure 7.4 illustrates the ancestral intersubunit bridges. Ancestral intersubunit bridges B3, B2b, and B2c are part of the early evolution of the SSU and LSU. Those bridges were formed within the first 3 phases of our evolutionary model. In SSU evolution, ancestral intersubunit bridge B3 (shown as orange in the Figure 7.4) is located in ancestral expansion segment 1. Ancestral expansion segment 2 governs the ancestral intersubunit bridge B2b and part of the ancestral intersubunit bridge B2c. Ultimately, ancestral expansion segment 3a and 5 formed the rest of the ancestral intersubunit bridge B2c. After those 3 bridges formed, the SSU and LSU were able to interact with each other. This process corresponds to phase 3 of our evolutionary model.

We propose that the ribosome at phase 4 is composed of associated subunits and capable of producing non-coded peptides by interacting with tRNA and proto-mRNA. The SSU requires the LSU to be stable in this phase. To this

end, the structure and stability of the SSU is dependent on the LSU. From this observation, we suggest that the SSU is younger than the LSU.

As for phase 5, we propose that the origin of coded peptide synthesis and translocation is started in this phase. Phase 5 is the transition phase for the ribosome from the primitive form to the more complex and functional form. In this phase, the ribosome was capable of producing complex proteins that led to production of more complex mechanisms and functions. Proteins start playing an important stability role in the ribosome after this phase. Since the tRNA anticodon stem loop gained the ability to interact with the ribosome, a non-coded catalytic ribozyme became a more complex and functional translocating machine. To this end, we can describe phase 5 as the most complex phase in SSU evolution.

In phase 6 of our evolutionary model, the common core of the rRNA matures. This phase mainly adds onto the surface of the ribosome and provides binding sites for ribosomal proteins for stability and structure. After completion of phase 6, the common core of the rRNA is completed and matured and becomes more stable and compact. Figure 7.5 shows the phase 6 ancestral expansions in red and clearly indicates the surface growth in this phase.

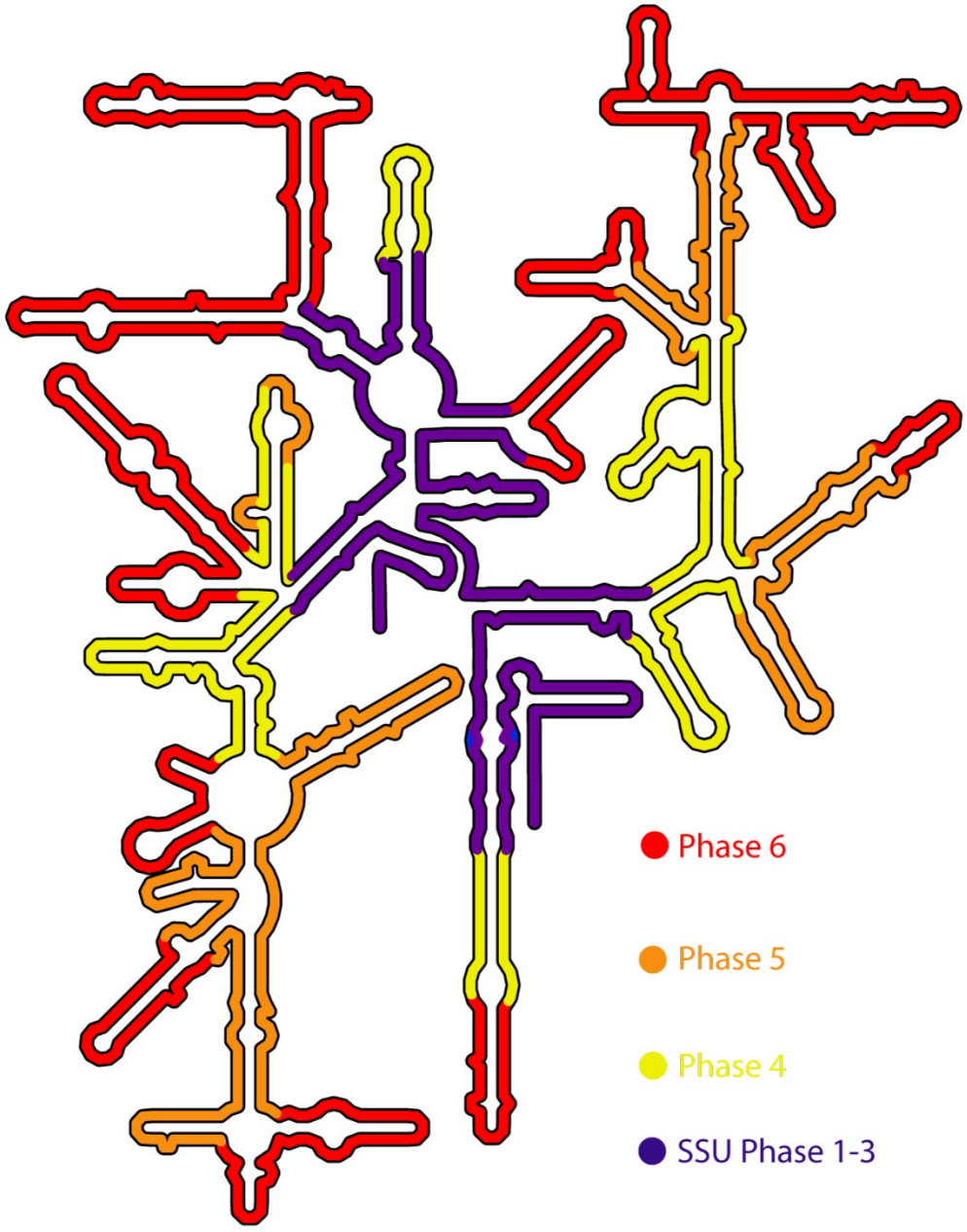

Figure 7.5 Phases of the SSU rRNA evolutionary model.
First 3 phases are shown as a combined evolutionary pathway in purple. The reason for this representation is the independent coevolution of the SSU and LSU through the first 3 phases. After phase 3, the SSU and LSU were capable of associating via ancestral intersubunit bridges. Phase 4 is shown as yellow. Phase 5 is shown as orange while phase 6 is shown as red. The secondary structure is from our in-house Ribovision server.

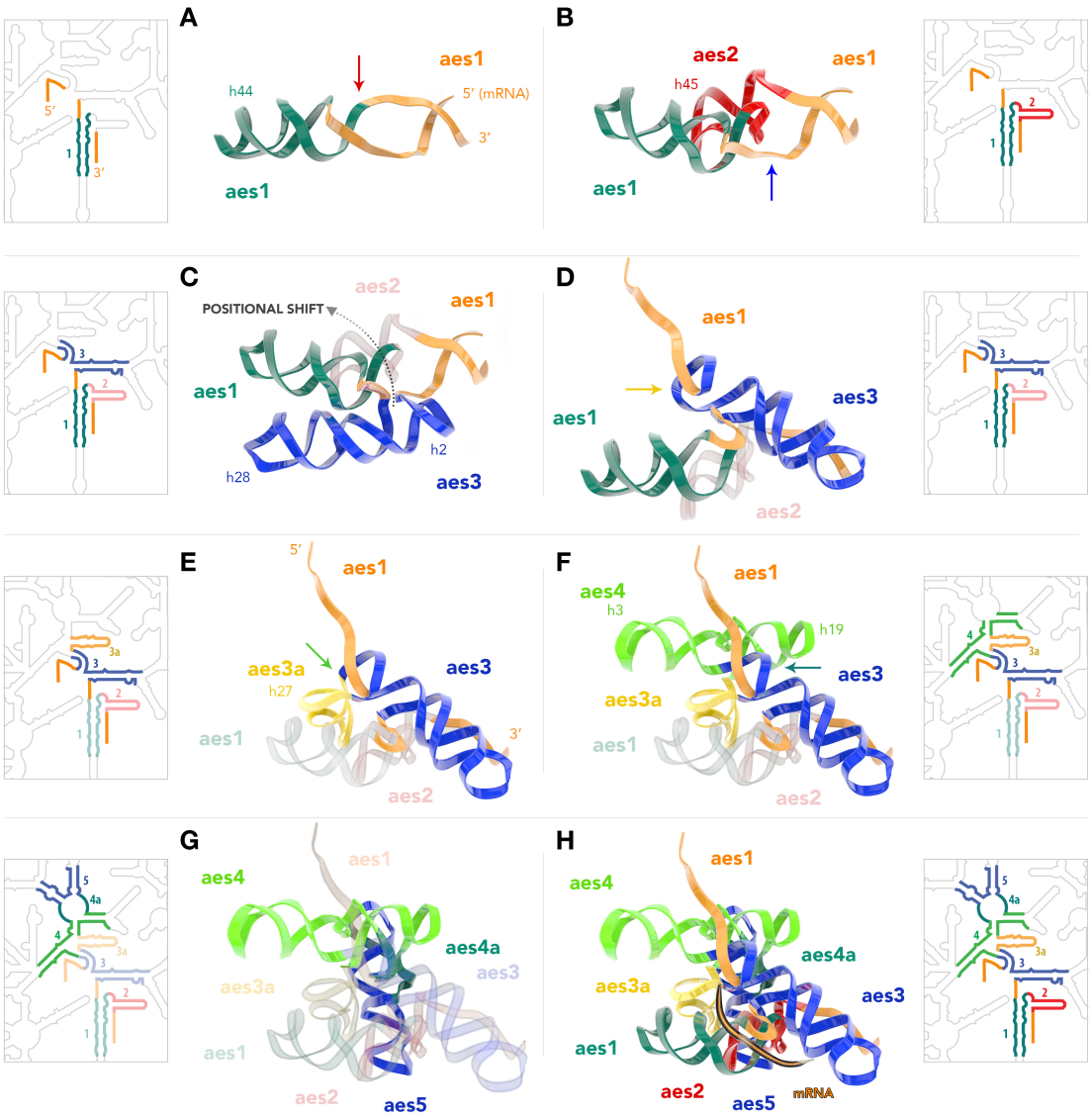

Figure 7.6 Origin and evolution of the central pseudoknot and decoding center of the SSU.
A) Aes1 forms a portion of helix 44 (teal) as well as the 3' and 5' ends of the SSU rRNA (orange). The 5' end of the rRNA is modeled by the mRNA of *T. thermophilus* (PDB ID: 2j00). B) Aes1 is expanded by insertion of aes 2 (red). C) Aes 3 (blue) is inserted at the 5' end at the opposite side from the aes2; D) Two insertions (aes 2 and aes 3) facing each other at the 3' and 5' ends clash and promote dissociation of the termini. The 5' end and aes 3 flip out. E) aes 3 is extended by aes 3a (yellow). F) aes4 (green) is inserted between aes 3 and 3a, forming a triple helix with the 5' termini, locking it into the central pseudoknot; G) aes 5 (blue) is acquired via aes 4a linkage (teal), providing a stability to the central pseudoknot; H) The remaining sticky 3' end associates with another single stranded RNA (via Shine-Dalgarno - anti-Shine Dalgarno helix) adding proto-mRNA into the association. The corresponding expansion segments mapped onto the secondary structure are shown in the insert panels.

7.3.2 Evolutionary model for the central pseudoknot and the central core

We propose that the single-stranded RNA binding functionality of SSU rRNA was the driving force for the early evolution of the SSU. It was previously proposed that the ancient functional RNA originated as a stem loop, in which the terminal ends are paired with each other (Petrov et al., 2014; Schimmel, 1991). Here we propose the paired termini phenomenon for the origins of SSU rRNA. Unlike the associated termini of the modern LSU, in early evolutional steps the SSU rRNA went a dissociation process for the termini. Dissociation of the terminal strands later led the entire evolutionary process to be distinct in terms of function and structure. The stem loop expands by acquiring expansion segments with unique insertion fingerprints.

Figure 7.6 summarizes and illustrates the origin of SSU evolution from a stem loop model. In detail, the stem loop here is the ancestral expansion segment 1 which governs the helix 44 and the 3' terminus of the SSU rRNA (the anti Shine-Dalgarno sequence is located in this terminus) and its complementary sequence as we mentioned earlier. Here we have modeled the complementary sequence from the mRNA, which holds the Shine-Dalgarno sequence (Figure 7.6A). After insertion of the ancestral expansion segments 2 and 3 as indicated by insertion fingerprints, a rearrangement occurs in the folding of five-way junction motif as a result of a more favorable conformational arrangement (Figure 7.6B-C). In this rearrangement process, strand termini dissociation occurs (Figure 7.6D). Furthermore, insertion of helix 27 extends the ancestral expansion segment 3 to the ancestral expansion segment 3a (Figure 7.6E). Insertion of the ancestral expansion segment 4 provides a docking site for the dissociated 5' terminus.

Acquisition of the ancestral expansion segment also leads to the formation of the central pseudoknot via triple helical interactions (Figure 7.6F). Stabilization of the central pseudoknot is enhanced after acquisition of the ancestral expansion segment 5 via the ancestral expansion segment 4a linkage (Figure 7.6G). The 3' terminus became unpaired and dissociated after completion of the early evolution model. By this connection, recognition of the mRNA through Shine-Dalgarno- anti Shine –Dalgarno interactions remained possible (Figure 7.6H). This model explains how the central core of SSU rRNA originated and how other domains of the SSU radiated from the central core in later steps of SSU evolution.

Our hypothesis is that central pseudoknot evolution was completed independent of the LSU evolution in that the LSU and SSU do not require each other to be functional until phase 3 of our evolutionary model. After establishment of the central core rRNA, association of SSU and LSU was driven by strand separation and single stranded RNA binding function. We also suggest that before the subunit association occurred, the LSU may have used different co-factors to interact with and stabilize the proto-tRNAs (Lehmann, 2000).

7.4 Conclusion

We have employed insertion fingerprints to track the evolutionary pathway of the SSU common core rRNA by temporal order of the ancestral expansion segments. We have superimposed rRNA three-dimensional structures and confirmed the integrity of the insertion fingerprints through eukaryotic expansion segments. By using this phenomenon, we have rooted back the SSU common core rRNA through the decoding center and proposed an evolutionary pathway through insertions by accretion. Our model explains the 3' and 5'

termini strand separation of SSU rRNA, which later provided binding space for mRNA. This is also the evolutionary history of the coding machinery.

Furthermore, we have proposed a timeline in which the SSU and LSU associated through ancestral intersubunit bridges. Molecular interactions between the SSU and LSU in early evolutionary steps supports the hypothesis that the SSU and LSU evolved independently from each other until phase 3 of our evolutionary model.

Ultimately, we have suggested the evolutionary steps for the central pseudoknot and the central core of the SSU rRNA. Originating from the decoding center, we have provided a detailed explanation of how the central pseudoknot evolved in the same manner of the insertion fingerprints. On the other hand, those results supports that the central core of the SSU rRNA was the ancient part of the small ribosomal subunit providing an anchor point for the other domains of the SSU by functioning as the structural core.

CHAPTER 8

CONCLUSIONS AND FUTURE DIRECTIONS

I have proposed revised secondary structure maps and a revised domain structure for the SSU rRNA (discussed in chapter 3 and chapter 4). Furthermore, I have designed an experimental system to test my domain A hypothesis and confirmed that domain A satisfies the criteria as a core domain (discussed in chapter 5). Further implications for the integrity of domain A are its folding and interactions with the ribosomal proteins (discussed in chapter 6). As a consequence, I have proposed an evolutionary model for the SSU and the central pseudoknot and domain A (discussed in chapter 7).

In this chapter, I summarize the results and discuss the future implications of the results. More importantly, I propose a research to help to elucidate the function of the central pseudoknot in translation as a future work.

8.1 More accurate secondary structures

We have proposed revised secondary structure maps for SSU rRNA from all domains of life. Historical secondary structure maps were predicted by co-variation analysis. The main problem with the co-variation analysis is the unpredicted non-canonical base pairs that cause misprediction of the central pseudoknot, the most significant structural core of the SSU rRNA, due to a lack of 3D information. Our approach involves using three-dimensional structures to observe interactions such as base-pairing, base stacking, base-phosphate, and base-sugar. We have tried to keep the conventional representation of the SSU

rRNA as much as possible and made our corrections based on the historical maps (Figure 3.1, Chapter 3).

Since our approach uses the experimentally resolved structural data, it is more accurate than co-variation analysis. Future implementation of the revised secondary structure maps will be assisting ribosome researchers to interpret their data more accurately. For example, one can consider non-canonical base pairs in the central pseudoknot and put this information into account rather than considering many nucleotides as non-pairing single strand which would cause a misinterpretation. Moreover, we have deposited revised secondary structure maps in our web server with a variety of data mapped on at (http://*apollo.chemistry.gatech.edu/RibosomeGallery*).

8.2 Revised domain structure

After correcting the secondary structure maps for the SSU rRNA, we proposed a revision of the domain architecture of the SSU rRNA. My research is motivated by the urge to improve the structural representation of the SSU rRNA. More specifically, the boundaries of the historical domains of the SSU rRNA are not well defined and the structure of the SSU rRNA is not properly reflected. We have defined the domain criteria by allocating each rRNA helix to a single domain and each nucleotide to a single helix. Moreover, each domain should have the ability to fold to its native structure if it is excised from the surrounding RNA. This process identifies a new structural core domain where all other domains radiate from it. We proposed this core domain as "domain A"(Figure 4.1, Chapter 4).

Domain A is sequentially and more significantly structurally highly conserved through different kingdoms of life. Our superimposition of three-dimensional structures from 5 different species gave an 0.778 Å RMSD with phosphates only, which is an indication of a structurally frozen core in the heart of the SSU rRNA (Figure 4.3, Chapter 4).

Furthermore, the core helices of domain A show mimicry of L-shaped structure of tRNAs, release factors etc. This is significant for making connections between early evolutionary events, assumptions of origins of the triple code and functional proto macromolecules (Figure 4.5, Chapter 4).

The future impact of the revised domain architecture is a better understanding of the structure and evolution of the ribosome and the dynamics of translocation. Well defined domain boundaries and an intact structure are significant for the overall progress of ribosome research. Additionally, our findings have implications for human diseases, drug targeting, ribosome biogenesis and molecular mimicry.

8.3 Domain A is confirmed as a domain

We have asked if domain A satisfies the criteria of autonomous folding. We have developed an experimental system to isolate the domain A from the rest of the SSU rRNA and characterize the domain A with methods such as SHAPE, CD spectroscopy, and truncation and mutation in the central pseudoknot. Our results confirm autonomous folding and ultimately domain A as a true domain (Chapter 5).

Our results support the hypothesis that the central pseudoknot is flexible and dynamic. Furthermore, we have shown that helix 28 of the SSU rRNA is

important for the proper folding of the central pseudoknot. Those results will help researchers to identify and isolate the central pseudoknot for studies such as drug screening, mutations, RNA folding, protein binding, and ribosome biogenesis.

8.4 Ribosomal Proteins S5 and S12

Since ribosomal proteins S5 and S12 are the only 30S proteins interacting the domain A nucleotides in the native 30S, we have tested their binding to our domain A^{ISO} model *in vivo* and *in vitro*. Our yeast-three hybrid assay results reveal that S5 and S12 are not binding to domain A^{ISO} RNA *in vivo*. These proteins contain many positively charged amino acid residues which may lead to an unspecific binding to nucleic acids. To this end, inside the yeast cells, these proteins can unspecifically interact with any nucleic acid to prevent specific binding to target domain A^{ISO} RNA.

Furthermore, we have shown that S5 fusion protein binds to different RNAs including domain A^{ISO}, P4-P6 domain of group I intron, and a 12 mer duplex RNA *in vitro* unspecifically. A ladder effect is observed in these interactions suggesting that multiple proteins bind to one RNA molecule.

Moreover, SHAPE and CD experiments reveal that binding of S5 fusion protein to domain A^{ISO} RNA causes a disruption in the tertiary structure of the domain A^{ISO} RNA, more specifically the central pseudoknot (Figure 6.9 and Figure 6.10, Chapter 6).

As we noted in chapter 6 , we hypothesize that positively charged residues of S5 are breaking up the tertiary structure of the central pseudoknot. Since the central pseudoknot is not very stable due to its flexibility, high positive

charge can cause disruption of the structure. In other words, folding of the central pseudoknot is affected by positively charged ribosomal protein S5. This hypothesis also explains why ribosomal protein S5 and also similarly charged S12 are not primary binding proteins in 30S assembly (Bunner et al., 2010; Mulder et al., 2010). Otherwise, primary binding of these proteins would retard the 16s rRNA folding by destabilizing the central pseudoknot in 30S assembly in that domain A acts as a hub which other domains radiate from. This hypothesis is also consistent with the fact that the domain A is buried in the core of 16S rRNA requiring other proteins bind to and stabilize the overall 16S structure prior to binding of highly basic proteins S5 and S12.

8.5 Evolutionary model for the SSU and the central pseudoknot

We have proposed an evolutionary model for the SSU rRNA and the central pseudoknot using insertion fingerprints. Insertion fingerprints are observed through eukaryotic expansion segments and confirmed by superimposition of different species. Our model elucidates how the SSU has grown into the modern subunit by originating from the decoding center. We call the accretion elements "ancestral expansion segments". We have divided this growth of the SSU rRNA into 6 phases by consistent molecular interactions and acquired functions. Temporal ordering of the ancestral expansion segments reveals dependency between the LSU and SSU after phase 3 of our evolutionary model (Figure 7.3, Chapter 7). Those results suggest that the evolution of the SSU and LSU was independent in the beginning.

Furthermore, our model suggests that the SSU went under strand separation through the 5' and 3' termini. Additionally, evolution of the central pseudoknot is clearly unearthed in our model (Figure 7.6, Chapter 7).

Overall, our SSU evolutionary model will have important implications on the study of the origin of early events in biology and biochemistry such as the origins of the genetic code, translocation, and coded and non-coded translation. Moreover, our model will be a useful tool for predicting the higher eukaryotic structures, which are very scarce recently.

8.6 Proposed research for solving the central pseudoknot mystery

The role of the central pseudoknot in translation is extremely significant however it is not yet fully understood. The experimental results from Brink and Poot (Brink et al., 1993; Poot et al., 1996) suggest that the formation of the central pseudoknot is vital for the subunit association. On the other hand, there are no direct contact residues located in the central pseudoknot, which associate with the LSU to form the ribosome assembly. This fact brings the idea to light that the central pseudoknot has an important structural role in SSU rRNA to facilitate its folding to the native state to allow the subunits to associate during translation. Our revised domain structure also supports this hypothesis in that the SSU domains radiate form a central hub, domain A, which governs the central pseudoknot. To this end, I propose future research to help elucidate the role of the central pseudoknot in ribosome function.

8.6.1 *In vivo* SHAPE for screening the changes in the 16S rRNA structure with defected central pseudoknot

SHAPE is a powerful method to screen RNA structures and especially changes in structure based on external or internal effects such as antibiotic binding or nucleotide mutations. I have discussed the C18A mutation and how this mutation inhibits the protein synthesis in chapter 2, section 2.2. Here I propose the SHAPE method *in vivo* to screen 16S rRNA in the presence and absence of the C18A mutation. This will allow one to compare mutant and wild type 16S rRNA and screen the changes in the entire 16S rRNA based on the mutations in the central pseudoknot. It has been recently shown that SHAPE can be used to screen ribosomal RNAs *in vivo* (McGinnis & Weeks, 2014).

8.6.2 How to purify 16S rRNA and isolate the mutant rRNA from chromosomally enchoded ribosomes.

I propose *E. coli* as a model organism for this research. *E. coli* has 7 operons to express ribosomal RNAs. In my proposed experimental system, one can express the mutant 16S rRNAs by using a plasmid governing the entire ribosomal operon rrnB (pK4-16). This low copy number plasmid contains a kanamycin marker and an *E. coli* rrnB operon. Transformation of this plasmid to a deletion strain (SQ141, this *E. coli* strain has 6 of its ribosomal operons deleted and it only has 1 chromosomal rRNA operon) will provide a homogenous ratio between chromosomally encoded ribosomes and mutant ribosomes.

My proposed experimental approach is simple. Once the plasmid pK4-16 (with the C18A mutation in 16S rRNA) is transformed to SQ141 deletion strains, cells can be grown up to mid-log phase in the presence of kanamycin. Cells

then can be treated with 5mM 1-methyl-7-nitroisatoic anhydride (1M7) in DMSO (final concentration) and incubated at 37 °C for 5 min by shaking. SHAPE reagent 1M7 is suitable for the *in vivo* reactions in that the nitro group at 7 position makes it fast reacting and it also penetrates into the cell (McGinnis & Weeks, 2014; Mortimer & Weeks, 2007). In parallel, a control background reaction can be performed with DMSO only. After the SHAPE reaction is complete, the cells can be harvested and total RNA can br purified by affinity chromatography.

8.6.2.1 How to differenciate the plasmid mediated 16S rRNA from chromosomally encoded 16S rRNA

Since total recovered RNA will include both mutant and wild type 16S rRNA, I propose an additional selective mutation in 16S rRNA. It has been shown previously that a mutation in Shine-Dalgarno sequence of the 16S rRNA does not perturb ribosome function or structure and renders a selective and active ribosome for specified proteins (Brink et al., 1993). Here I propose a serial mutation in helix 45 of *E. coli* 16S rRNA which is structurally in close proximity to the anti-Shine –Dalgarno sequence. A total 7 nucleotide mutations out of 20 nucleotides in helix 45 can provide high selectivity for mutant ribosomes without perturbing the structure and function. Those mutations are not selected randomly. Those mutations correspond to the deviation in helix 45 of *E. coli* and occur in another bacteria, *T. thermophilus (5* nucleotides) and a eukarya, *S. cerevisiae* (2 nucleotides). Figure 8.1 demonstrates the modification in helix 45 and C18A. There are a total of 7 nucleotide deviations in helix 45 in *E. coli* compared to *T. thermophilus*. It has been shown that *E. coli* can survive with *T.*

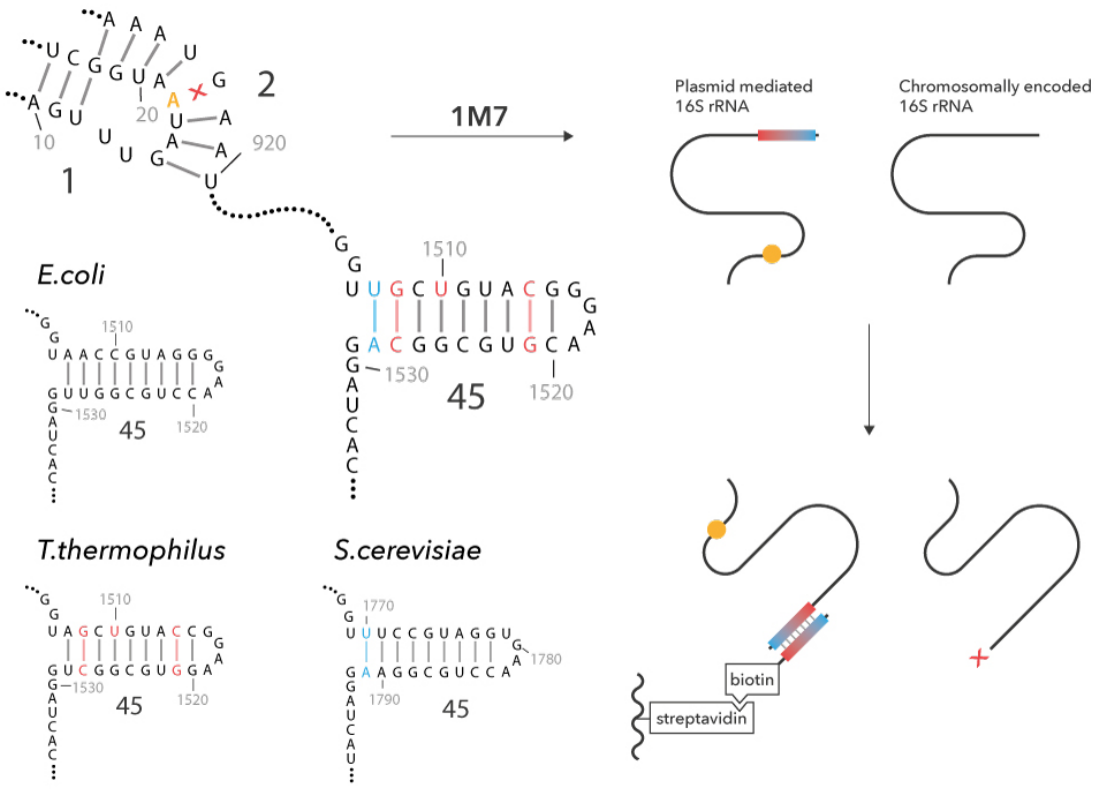

Figure 8.1 Modifications in helix 45 of *E. coli* 16S rRNA.
Wild type helix 45 from E. coli 16S is seen on the left. Mutations corresponding to *T. thermophilus* helix 45 is shown as red while mutaions corresponding to S. cerevisiae is shown as blue. Affinity purification with a streptavidin column is demonstrated on the right. The blue- red rectangle on the 16S rRNA represents the modified helix 45, while the yellow dot represents the C18A mutation. Biotin labelled helix 45 complementary oligonucleotide is represented on the right bottom side. Chromosomally encoded 16S rRNA will not have affinity to the column due to lack of complementary sequence.

thermophilus 16S rRNA replacement ((Thompson & Dahlberg, 2004)). In other words, when *T. thermophilus* 16S rRNA sequence is inserted to E. coli as the SSU rRNA, *E. coli* can maintain its cellular activity. To this end, 7 nucleotide mutations in 16S rRNA should be well compensated and not affect the 16S rRNA structure once helix 45 is mutated. Overall, this change in 16S rRNA will make the C18A mutant ribosomes so much more favorable to the oligonucleotide attached affinity column than chromosomally encoded wild type ribosomes.

8.6.3 Cappilary electrophoresis and data processing

Once the SHAPE modified mutant SSU rRNA is obtained and purified, capillary electrophoresis can follow to determine the modified residues. For this purpose, four 5' – FAM modified primers can be used in a reverse transcription reaction to convert the modified data from RNA to complementary DNA by primer extension. The results can be processed as previously described (Chapter 5 section 5.2.4). Changes in 16S rRNA governing the mutant central pseudoknot compared to wild type might provide valuable insight into the mechanism of inhibition of protein synthesis via this vital mutation. Furthermore, those results might also provide significant clues about the decoding and translocation process during protein synthesis. I believe that this research can bring a major impact in terms of the small ribosomal subunit and its function in translation.

REFERENCES

Ada, G., & Perry, B. T. (1954). The nucleic acid content of influenza virus. *The Australian journal of experimental biology and medical science, 32*(4), 453-468.

Anger, A. M., Armache, J. P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D. N., & Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. *Nature, 497*(7447), 80-85. doi: 10.1038/nature12104

Armache, J.-P., Anger, A. M., Márquez, V., Franckenberg, S., Fröhlich, T., Villa, E., . . . Wilson, D. N. (2013). Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Research, 41*(2), 1284-1293. doi: 10.1093/nar/gks1259

Armache, J. P., Jarasch, A., Anger, A. M., Villa, E., Becker, T., Bhushan, S., . . . Beckmann, R. (2010). Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5 Å resolution. *Proc Natl Acad Sci U S A, 107*(46), 19748-19753. doi: 10.1073/pnas.1009999107

Athavale, S. S., Gossett, J. J., Hsiao, C., Bowman, J. C., O'Neill, E., Hershkovitz, E., . . . Williams, L. D. (2012). Domain III of the T. thermophilus 23S rRNA folds independently to a near-native state. *RNA (New York, N.Y.), 18*(4), 752-758. doi: 10.1261/rna.030692.111

Athavale, S. S., Petrov, A. S., Hsiao, C., Watkins, D., Prickett, C. D., Gossett, J. J., . . . Williams, L. D. (2012). RNA folding and catalysis mediated by iron (II). *PloS one, 7*(5), e38024-e38024. doi: 10.1371/journal.pone.0038024

Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science, 289*(5481), 905-920.

Belanger, F., Theberge-Julien, G., Cunningham, P. R., & Brakier-Gingras, L. (2005). A functional relationship between helix 1 and the 900 tetraloop of 16S ribosomal RNA within the bacterial ribosome. *Rna, 11*(6), 906-913. doi: 10.1261/rna.2160405

Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., & Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 A resolution. *Science, 334*(6062), 1524-1529. doi: 10.1126/science.1212642

Ben-Shem, A., Jenner, L., Yusupova, G., & Yusupov, M. (2010). Crystal structure of the eukaryotic ribosome. *Science, 330*(6008), 1203-1209. doi: 10.1126/science.1194294

Bendak, K., Loughlin, F., Cheung, V., O'Connell, M., Crossley, M., & Mackay, J. (2012). A rapid method for assessing the RNA-binding potential of a protein. *Nucleic acids research*, gks285.

Bernier, C., Petrov, A. S., Waterbury, C., Jett, J., Li, F., Freil, L. E., . . . Williams, L. D. (2014). RiboVision: Visualization and Analysis of Ribosomes. *Faraday Discussions, 169*(1), 1–12. doi: 10.1039/C3FD00126A

Bokov, K., & Steinberg, S. V. (2009). A hierarchical model for evolution of 23S ribosomal RNA. *Nature, 457*(7232), 977-980. doi: 10.1038/nature07749

Bowman, J. C., Lenz, T. K., Hud, N. V., & Williams, L. D. (2012). Cations in charge : magnesium ions in RNA folding and catalysis. *Current opinion in structural ….*

Brink, M. F., Verbeet, M. P., & de Boer, H. A. (1993). The central pseudoknot connecting the three major domains in 16S rRNA is required for translational initiation *The Translational Apparatus* (pp. 209-220): Springer.

Brink, M. F., Verbeet, M. P., & de Boer, H. a. (1993). Formation of the central pseudoknot in 16S rRNA is essential for initiation of translation. *The EMBO journal, 12*(10), 3987-3996.

Bunner, A. E., Beck, A. H., & Williamson, J. R. (2010). Kinetic cooperativity in Escherichia coli 30S ribosomal subunit reconstitution reveals additional complexity in the assembly landscape. *Proceedings of the National Academy of Sciences, 107*(12), 5417-5422.

Butcher, S. E., & Pyle, A. M. (2011). The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. *Accounts of Chemical Research, 44*(12), 1302-1311. doi: 10.1021/ar200098t

Bycroft, M., Grunert, S., Murzin, A. G., Proctor, M., & St Johnston, D. (1995). NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J, 14*(14), 3563-3571.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., . . . Gutell, R. R. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC bioinformatics, 3*, 2-2.

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature, 407*(6802), 340-348.

Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N., & Noller, H. F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science, 285*(5436), 2095-2104. doi: 10.1126/science.285.5436.2095

Chao, F.-C., & Schachman, H. (1956). The isolation and characterization of a macromolecular ribonucleoprotein from yeast. *Archives of biochemistry and biophysics, 61*(1), 220-230.

Chen, X., Sim, S., Wurtmann, E. J., Feke, A., & Wolin, S. L. (2014). Bacterial noncoding Y RNAs are widespread and mimic tRNAs. *Rna*.

Clark, C. G., Tague, B. W., Ware, V. C., & Gerbi, S. A. (1984). Xenopus laevis 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. *Nucleic Acids Research, 12*(15), 6197-6220.

Deigan, K. E., Li, T. W., Mathews, D. H., & Weeks, K. M. (2009). Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci U S A, 106*(1), 97-102. doi: 10.1073/pnas.0806929106

Doudna, J. A., & Cech, T. R. (2002). The chemical repertoire of natural ribozymes. *Nature, 418*(6894), 222-228.

Draper, D. E., & Reynaldo, L. P. (1999). RNA binding strategies of ribosomal proteins. *Nucleic Acids Res, 27*(2), 381-388.

Dunkle, J. A., Wang, L. Y., Feldman, M. B., Pulk, A., Chen, V. B., Kapral, G. J., . . . Cate, J. H. D. (2011b). Structures of the Bacterial Ribosome in Classical and Hybrid States of tRNA Binding. *Science, 332*(6032), 981-984. doi: 10.1126/science.1202692

Felden, B., Florentz, C., Giegé, R., & Westhof, E. (1996). A central pseudoknotted three-way junction imposes tRNA-like mimicry and the orientation of three 5'upstream pseudoknots in the 3'terminus of tobacco mosaic virus RNA. *Rna, 2*(3), 201.

Fitch, W. M. (1971). Toward defining course of evolution - minimum change for a specific tree topology. *Systematic Zoology, 20*(4), 406-&. doi: 10.2307/2412116

Fox, G. E., & Woese, C. R. (1975). 5S RNA secondary structure. *Nature, 256*(5517), 505-507.

Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., . . . Agrawal, R. K. (1995). A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature, 376*(6539), 441-444. doi: 10.1038/376441a0

Frederiksen, J. K., Li, N.-S., Das, R., Herschlag, D., & Piccirilli, J. A. (2012). Metal-ion rescue revisited: Biochemical detection of site-bound metal ions important for RNA folding. *Rna, 18*(6), 1123-1141.

Fujiwara, T., Ito, K., & Nakamura, Y. (2001). Functional mapping of ribosome-contact sites in the ribosome recycling factor: a structural view from a tRNA mimic. *Rna, 7*(1), 64-70.

Fukushi, S., Okada, M., Stahl, J., Kageyama, T., Hoshino, F. B., & Katayama, K. (2001). Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. *J Biol Chem, 276*(24), 20824-20826. doi: 10.1074/jbc.C100206200

Gabashvili, I. S., Agrawal, R. K., Grassucci, R., Squires, C. L., Dahlberg, A. E., & Frank, J. (1999). Major rearrangements in the 70S ribosomal 3D structure caused by a conformational switch in 16S ribosomal RNA. *The EMBO Journal, 18*(22), 6501-6507. doi: 10.1093/emboj/18.22.6501

Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S. M., . . . Frank, J. (2003). Study of the Structural Dynamics of the E. coli 70S Ribosome Using Real-Space Refinement. *Cell, 113*(6), 789-801. doi: 10.1016/s0092-8674(03)00427-6

Gao, N., & Frank, J. (2006). A library of RNA bridges. *Nature chemical biology, 2*(5), 231-232.

Gerbi, S. A. (1996). Expansion segments: Regions of variable size that interrupt the universal core secondary structure of ribosomal RNA. In R. A. Zimmermann & A. E. Dahlberg (Eds.), *Ribosomal RNA—Structure, evolution, processing, and function in protein synthesis* (pp. 71–87). Boca Raton, FL: CRC Press.

Giegé, R., Frugier, M., & Rudinger, J. (1998). tRNA mimics. *Current opinion in structural biology, 8*(3), 286-293. doi: http://dx.doi.org/10.1016/S0959-440X(98)80060-2

Gregory, S. T., & Dahlberg, A. E. (2009). Genetic and structural analysis of base substitutions in the central pseudoknot of *Thermus thermophilus* 16S ribosomal RNA. *RNA, 15*(2), 215-223. doi: 10.1261/rna.1374809

Guo, Z., & Noller, H. F. (2012). Rotation of the head of the 30S ribosomal subunit during mRNA translocation. *Proc Natl Acad Sci U S A, 109*(50), 20391-20394. doi: 10.1073/pnas.1218999109

Gutell, R. R., Larsen, N., & Woese, C. R. (1994). Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev, 58*(1), 10-26.

Gutell, R. R., Lee, J. C., & Cannone, J. J. (2002). The accuracy of ribosomal RNA comparative structure models. *Current opinion in structural biology, 12*(3), 301-310.

Hajdin, C. E., Bellaousov, S., Huggins, W., Leonard, C. W., Mathews, D. H., & Weeks, K. M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences of the United States of America, 110*(14), 5498-5503. doi: 10.1073/pnas.1219988110

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., . . . Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell, 107*(5), 679-688.

Hassouna, N., Michot, B., & Bachellerie, J. P. (1984). The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucleic Acids Research, 12*(8), 3563-3583.

Hirokawa, G., Kiel, M. C., Muto, A., & Selmer, M. (2002). Post-termination complex disassembly by ribosome recycling factor, a functional tRNA mimic. *The EMBO …, 21*(9), 2272-2281.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., . . . Zamir, A. (1965). Structure of a ribonucleic acid. *Science, 147*(3664), 1462-1465.

Hsiao, C., Lenz, T. K., Peters, J. K., Fang, P.-Y., Schneider, D. M., Anderson, E. J., . . . Dean Williams, L. (2013). Molecular paleontology: a biochemical model of the ancestral ribosome. *Nucleic acids research*(Figure 1), 1-13. doi: 10.1093/nar/gkt023

Hsiao, C., Mohan, S., Kalahar, B. K., & Williams, L. D. (2009). Peeling the onion: ribosomes are ancient molecular fossils. *Molecular Biology and Evolution, 26*(11), 2415-2425. doi: 10.1093/molbev/msp163

Johnson, G. T., Autin, L., Goodsell, D. S., Sanner, M. F., & Olson, A. J. (2011). ePMV Embeds Molecular Modeling into Professional Animation Software Environments. *Structure, 19*(3), 293-303. doi: Doi 10.1016/J.Str.2010.12.023

Juzumiene, D. I., & Wollenzien, P. (2001). Arrangement of the central pseudoknot region of 16S rRNA in the 30S ribosomal subunit determined by site-directed 4-thiouridine crosslinking. *Rna, 7*(1), 71-84.

Kim, S., Quigley, G., Suddath, F., McPherson, A., Sneden, D., Kim, J., . . . Rich, A. (1973). Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science, 179*(4070), 285-288.

Ladner, J., Jack, A., Robertus, J., Brown, R., Rhodes, D., Clark, B., & Klug, A. (1975). Structure of yeast phenylalanine transfer RNA at 2.5 A resolution. *Proceedings of the National Academy of Sciences, 72*(11), 4414-4418.

Lata, K. R., Agrawal, R. K., Penczek, P., Grassucci, R., Zhu, J., & Frank, J. (1996). Three-dimensional reconstruction of the Escherichia coli 30 S ribosomal subunit in ice. *J Mol Biol, 262*(1), 43-52. doi: 10.1006/jmbi.1996.0497

Lehmann, J. (2000). Physico-chemical constraints connected with the coding properties of the genetic system. *Journal of Theoretical Biology, 202*(2), 129-144. doi: 10.1006/jtbi.1999.1045

Leontis, N. B., Stombaugh, J., & Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research, 30*(16), 3497-3531.

Leontis, N. B., & Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *Journal of molecular biology, 283*(3), 571-583.

Lescoute, A., & Westhof, E. (2006). The interaction networks of structured RNAs. *Nucleic Acids Research, 34*(22), 6587-6604. doi: 10.1093/nar/gkl963

Lodmell, J. S., & Dahlberg, A. E. (1997). A conformational switch in Escherichia coli 16S ribosomal RNA during decoding of messenger RNA. *Science, 277*(5330), 1262-1267.

Lund, J. M., Alexopoulou, L., Sato, A., Karow, M., Adams, N. C., Gale, N. W., . . . Flavell, R. A. (2004). Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proceedings of the National Academy of Sciences of the United States of America, 101*(15), 5598-5603.

Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A., & Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Research, 29*(22), 4724-4735.

Maris, C., Dominguez, C., & Allain, F. H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J, 272*(9), 2118-2131. doi: 10.1111/j.1742-4658.2005.04653.x

McGinnis, J. L., & Weeks, K. M. (2014). Ribosome RNA assembly intermediates visualized in living cells. *Biochemistry, 53*(19), 3237-3247. doi: 10.1021/bi500198b

Mears, J. A., Cannone, J. J., Stagg, S. M., Gutell, R. R., Agrawal, R. K., & Harvey, S. C. (2002). Modeling a minimal ribosome based on comparative sequence analysis. *Journal of Molecular Biology, 321*(2), 215-234.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., & Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nature Structural & Molecular Biology, 19*(6), 560-567. doi: 10.1038/nsmb.2313

Merino, E. J., Wilkinson, K. A., Coughlan, J. L., & Weeks, K. M. (2005). RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *Journal of the American Chemical Society, 127*, 4223-4231.

Michot, B., Qu, L.-H., & Bachellerie, J.-P. (1990). Evolution of large-subunit rRNA structure. *European Journal of Biochemistry, 188*(2), 219-229. doi: 10.1111/j.1432-1033.1990.tb15393.x

Mizushima, S., & Nomura, M. (1970). Assembly mapping of 30S ribosomal proteins from E. coli. *Nature, 226*, 1214-1218.

Mortimer, S. a., & Weeks, K. M. (2007). A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *Journal of the American Chemical Society, 129*(14), 4144-4145. doi: 10.1021/ja0704028

Mortimer, S. A., & Weeks, K. M. (2008). Time-Resolved RNA SHAPE Chemistry. *Journal of the American Chemical Society, 130*, 16178-16180.

Mulder, A. M., Yoshioka, C., Beck, A. H., Bunner, A. E., Milligan, R. A., Potter, C. S., . . . Williamson, J. R. (2010). Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit. *Science, 330*(6004), 673-677.

Muth, G. W., Ortoleva-Donnelly, L., & Strobel, S. A. (2000). A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science, 289*(5481), 947-950.

Nissen, P., Hansen, J., Ban, N., Moore, P. B., & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science, 289*(5481), 920-930.

Noeske, J., & Cate, J. H. (2012). Structural basis for protein synthesis: snapshots of the ribosome in motion. *Curr Opin Struct Biol, 22*(6), 743-749. doi: 10.1016/j.sbi.2012.07.011

Noller, H. F. (2005). RNA Structure: Reading the Ribosome. *Science, 309*(5740), 1508-1514.

Noller, H. F., Hoffarth, V., & Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science, 256*(5062), 1416-1419.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., . . . Waese, C. R. (1981). Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res, 9*(22), 6167-6189.

Noller, H. F., Kop, J., Wheaton, V., Brosius, J., Gutell, R. R., Kopylov, A. M., . . . Woese, C. R. (1981). Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Research, 9*(22), 6167-6189.

Noller, H. F., & Woese, C. R. (1981). Secondary structure of 16S ribosomal RNA. *Science, 212*(4493), 403-411.

Ogle, J. M., Brodersen, D. E., Clemons, W. M., Tarry, M. J., Carter, a. P., & Ramakrishnan, V. (2001). Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science (New York, N.Y.), 292*(5518), 897-902. doi: 10.1126/science.1060612

Petrov, A. S., Bernier, C. R., Gulen, B., Waterbury, C. C., Hershkovits, E., Hsiao, C. L., . . . Williams, L. D. (2014). Secondary Structures of rRNAs from All Three Domains of Life. *PloS one, 9*(2). doi: 10.1371/journal.pone.0088222

Petrov, A. S., Bernier, C. R., Hershkovits, E., Xue, Y., Waterbury, C. C., Hsiao, C., . . . Williams, L. D. (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Res, 41*(15), 7522-7535. doi: 10.1093/nar/gkt513

Petrov, A. S., Bernier, C. R., Hsiao, C., Norris, A. M., Kovacs, N. A., Waterbury, C. C., . . . Williams, L. D. (2014). Evolution of the ribosome at atomic resolution. *Proc Natl Acad Sci U S A, 111*(28), 10251-10256. doi: 10.1073/pnas.1407205111

Pleij, C. W., Rietveld, K., & Bosch, L. (1985). A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Research, 13*(5), 1717-1731.

Poot, R. A., Jeeninga, R. E., Pleij, C. W. A., & van Duin, J. (1997). Acetylation of ribosomal protein S5 affected by defects in the central pseudoknot in 16S ribosomal RNA? *FEBS letters, 401*(2-3), 175-179. doi: 10.1016/s0014-5793(96)01467-6

Poot, R. a., Pleij, C. W., & van Duin, J. (1996). The central pseudoknot in 16S ribosomal RNA is needed for ribosome stability but is not essential for 30S initiation complex formation. *Nucleic acids research, 24*(19), 3670-3676.

Poot, R. a., van den Worm, S. H., Pleij, C. W., & van Duin, J. (1998). Base complementarity in helix 2 of the central pseudoknot in 16S rRNA is essential for ribosome functioning. *Nucleic acids research, 26*(2), 549-553.

Powers, T., & Noller, H. F. (1991). A functional pseudoknot in 16S ribosomal RNA. *The EMBO journal, 10*(8), 2203-2214.

Quigley, G. J., & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science, 194*(4267), 796-806.

Ramakrishnan, V., & White, S. W. (1992). The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature, 358*(6389), 768-771.

Ramakrishnan, V., & White, S. W. (1998). Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. *Trends in biochemical sciences, 23*(6), 208-212.

Ramaswamy, P., & Woodson, S. A. (2009). S16 throws a conformational switch during assembly of 30S 5' domain. *Nat Struct Mol Biol, 16*(4), 438-445. doi: 10.1038/nsmb.1585

Reza, F. M. (1994). *An introduction to information theory*: Courier Corporation.

Rich, A., & RajBhandary, U. (1976). Transfer RNA: molecular structure, sequence, and properties. *Annual Review of Biochemistry, 45*(1), 805-860.

Richards, E. G. (1969). 5S RNA. An analysis of possible base pairing schemes. *European Journal of Biochemistry, 10*(1), 36-42.

Rivas, E., & Eddy, S. R. (2000). The language of RNA: a formal grammar that includes pseudoknots. *Bioinformatics, 16*(4), 334-340. doi: 10.1093/bioinformatics/16.4.334

Römer, R., & Hach, R. (1975). tRNA conformation and magnesium binding. *European Journal of Biochemistry, 55*(1), 271-284.

Sarver, M., Zirbel, C. L., & Stombaugh, J. (2008). FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *Journal of mathematical …, 56*(2000), 215-252. doi: 10.1007/s00285-007-0110-x.FR3D

Schimmel, P. (1991). RNA minihelices and the decoding of genetic information. *FASEB Journal, 5*(8), 2180-2187.

Schmeing, T. M., Huang, K. S., Strobel, S. A., & Steitz, T. A. (2005). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature, 438*(7067), 520-524.

Schrodinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.3r1.

Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., & Cate, J. H. D. (2005). Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science, 310*(5749), 827-834. doi: 10.1126/science.1117230

Searls, D. B. (1992). The Linguistics of DNA. *American Scientist, 80*(6), 579-591.

Segerstolpe, A., Granneman, S., Bjork, P., de Lima Alves, F., Rappsilber, J., Andersson, C., . . . Wieslander, L. (2013). Multiple RNA interactions position Mrd1 at the site of the small subunit pseudoknot within the 90S pre-ribosome. *Nucleic Acids Research, 41*(2), 1178-1190. doi: 10.1093/nar/gks1129

Selmer, M., Dunham, C. M., Murphy, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., . . . Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science (New York, N.Y.), 313*(5795), 1935-1942. doi: 10.1126/science.1131127

Serra, M. J., Baird, J. D., Dale, T., Fey, B. L., Retatagos, K., & Westhof, E. (2002). Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *Rna, 8*(03), 307-323.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3-55.

Shine, J., & Dalgarno, L. (1974). The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Robosome Binding Site. … *of the National Academy of Sciences, 71*(4), 1342-1346.

Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A., & Weeks, K. M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods, 11*(9), 959-965. doi: 10.1038/nmeth.3029

Sievers, A., Beringer, M., Rodnina, M. V., & Wolfenden, R. (2004). The ribosome as an entropy trap. *Proceedings of the National Academy of Sciences of the United States of America, 101*(21), 7897-7901.

Sim, A. Y., & Levitt, M. (2011). Clustering to identify RNA conformations constrained by secondary structure. *Proc Natl Acad Sci U S A, 108*(9), 3590-3595. doi: 10.1073/pnas.1018653108

Smit, S., Rother, K., Heringa, J., & Knight, R. (2008). From knotted to nested RNA structures: a variety of computational methods for pseudoknot removal. *RNA, 14*(3), 410-416. doi: 10.1261/rna.881308

Soding, J., & Lupas, A. N. (2003). More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays, 25*(9), 837-846. doi: 10.1002/bies.10321

Stelzl, U., Connell, S., Nierhaus, K. H., & Wittmann-Liebold, B. (2001). Ribosomal proteins: role in ribosomal functions. *eLS*.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification, 41*(1), 207-234.

Thompson, J., & Dahlberg, A. E. (2004). Testing the conservation of the translational machinery over evolution in diverse environments: assaying Thermus thermophilus ribosomes and initiation factors in a coupled transcription–translation system from Escherichia coli. *Nucleic acids research, 32*(19), 5954-5961.

Thompson, J., Kim, D. F., O'Connor, M., Lieberman, K. R., Bayfield, M. A., Gregory, S. T., . . . Dahlberg, A. E. (2001). Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proceedings of the National Academy of Sciences, 98*(16), 9002-9007.

Valle, M., Zavialov, A., Sengupta, J., Rawat, U., Ehrenberg, M., & Frank, J. (2003). Locking and Unlocking of Ribosomal Motions. *Cell, 114*(1), 123-134. doi: http://dx.doi.org/10.1016/S0092-8674(03)00476-8

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in biochemical sciences, 24*(11), 437-440.

Waterman, M. S., & Smith, T. F. (1978). Rna secondary structure - complete mathematical-analysis. *Mathematical Biosciences, 42*(3-4), 257-266. doi: 10.1016/0025-5564(78)90099-8

Weeks, K. M. (2010). Advances in RNA Secondary and Tertiary Structure Analysis by Chemical Probing. *Current opinion in structural biology, 20*(3), 295-304. doi: 10.1016/j.sbi.2010.04.001.Advances

Weijia, X., Wongsa, A., Jung, L., Lei, S., Cannone, J. J., & Gutell, R. R. (2011). RNA2DMap: A Visual Exploration Tool of the Information in RNA's Higher-Order Structure *Proceedings of 2011 IEEE International Conference on Bioinformatics and Biomedicine* (pp. 613-617). Atlanta, GA.

Wilkinson, K. a., Merino, E. J., & Weeks, K. M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature protocols, 1*(3), 1610-1616. doi: 10.1038/nprot.2006.249

Wilson, D. N., & Nierhaus, K. H. (2009). Ribosomal proteins: role in ribosomal functions. *eLS*.

Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., . . . Noller, H. F. (1980b). Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res, 8*(10), 2275-2293.

Woodson, S. A. (2010). Compact intermediates in RNA folding. *Annual review of biophysics, 39*, 61-77.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research, 31*(13), 3406-3415.